Gelsight Robotic Fingertip

by

Melody Grace Liu

Submitted to the Department of Mechanical Engineering
in partial fulfillment of the requirements for the degree of

Bachelor of Science in Mechanical Engineering

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

February 2017

© Massachusetts Institute of Technology 2017. All rights reserved.

Signature redacted

Author ............................

Department of Mechanical Engineering

January 20, 2017

Signature redacted

Certified by ............................................

Alberto Rodriguez
Assistant Professor
Thesis Supervisor

Signature redacted

Accepted by .............................................

Rohit Karnik
Professor of Mechanical Engineer, Undergraduate Officer
DISCLAIMER NOTICE

Due to the condition of the original material, there are unavoidable flaws in this reproduction. We have made every effort possible to provide you with the best copy available.

Thank you.

The images contained in this document are of the best quality available.
Gelsight Robotic Fingertip

by

Melody Grace Liu

Submitted to the Department of Mechanical Engineering
on January 20, 2017, in partial fulfillment of the
requirements for the degree of
Bachelor of Science in Mechanical Engineering

Abstract

Tactile sensing is integral in robotic manipulation, enabling robot hands to identify
objects and explore the environment. The gelsight tactile sensor designed in MIT
CSAIL is able to reconstruct the surface topography of objects up to submicron
accuracy. This thesis is focused on the integration of the gelsight tactile sensor into
the MIT Team’s gripper for the 2017 Amazon Robotics Challenge.

The design of the sensor is dependent on the three manipulation primitives in the
gripper platform: scooping, grasping, and suction. This paper discusses the primi-
tives and the sensory inputs they would require to create more precise manipulation
models. We then propose a gelsight tactile sensor around those primitives, creating a
design that can recover shear, deflection, compliance, and contact area information.
Lastly, we implement some of the image processing to recover deflection by creating a
mapping between image points and ground truth vicon values, producing an 75% ac-
curacy rate within a distance threshold, a value that can be improved with increased
measurements and better image processing techniques.

Further work focuses on image processing for shear, compliance, and contact area,
to be integrated into the overall design for the 2017 ARC gripper.

Thesis Supervisor: Alberto Rodriguez
Title: Assistant Professor
Acknowledgments

This thesis work wouldn't have been possible without the support and input of people who have helped me. I'm grateful for Professor Alberto Rodriguez, for taking me into the lab and allowing me to experiment with things I had never worked on before. Many thanks to Elliott Donlan for being a fantastic mentor and a great person to work with, and with whom this work would not have been possible. Thank you to Siyuan Dong who helped with data collection and the image processing. To Ted Adelson and Wenzhen Yuan, thank you for your help and technical advice in computer vision and gelsight. Lastly, thank you to my parents, family, and friends, who have graciously supported me through my academic career at MIT.
Contents

1 Background 9
  1.1 Motivation: Amazon Robotics Challenge 9
  1.2 Description of tactile sensing 10
  1.3 Gelsight 11
  1.4 Statement of Purpose 12

2 Development 13
  2.1 Finger Design 13
  2.2 Gelsight use in Manipulation Primitives 14
    2.2.1 Grasping 15
    2.2.2 Scooping 16
    2.2.3 Suction 17
    2.2.4 Exploration 18

3 Proposed Approach 19
  3.1 Manufacture of fingertips 19
  3.2 Manufacture of gelsight 20
  3.3 Electrical Setup 21

4 Vision processing to recover deflection data 23
  4.1 Experimental Setup 24

5 Conclusion 27
List of Figures

1-1 End-effector for Team MIT’s 2016 Amazon Robotics Competition entry. Two fingers with compliant fingertips are mounted on a parallel jaw gripper to grasp, scoop, and suction items [2] ........................................ 10

1-2 Uses of touch sensing in manipulation. Each sensor is shown on the left, linked to the sensor data they output. [3] ............................... 11

1-3 (a) A cookie is pressed against the skin of a gelsight block. (b) The skin is distorted, as shown in this view from beneath. (c) The cookie’s shape can be measured using photometric stereo and rendered at a novel viewpoint. [5] ................................................................. 12

2-1 Placement of gelsight on ARC end-effector and illustration of fingertip spatulas ................................................................. 14

2-2 This is horizontal grasping, where friction between the fingers and object surface prevent the object from slipping. (a) Smaller objects can be grasped at the fingertips. (b) Larger objects provide more surface area for grasping [2] ................................................................. 15

2-3 Side view of scooping primitive. The gripper preloads the fingernail to slide it underneath the object. Then, it slides forward until the object pushes against the back wall and slides higher onto the finger body. [2] 16

2-4 Side view of suction primitive ........................................... 17

3-1 The camera and mirror placement to recover gelsight image ....... 19
3-2 We show the gelsight finger covered in specular gold paint and a grid of small dots for tracking shear. An edge contact area feature can be seen.

3-3 We use 6 large checkerboard markers to recover fingertip deflection.

3-4 Electrical system to communicate between raspberry pi and computer

4-1 On the left, we can see the overall setup used to create the mapping. On the right, we see a closeup of the vicon markers.

4-2 On the left is the original image, with centroid points overlaid in red. On the right is the image after thresholding.

4-3 Deflection mapping setup on a more accurate fingertip shape.
Chapter 1

Background

1.1 Motivation: Amazon Robotics Challenge

The Amazon Robotics Challenge [1] is an annual competition that focuses on the unstructured problem of autonomously picking and stowing objects stored on shelves. Teams design an autonomous system that can stow objects into a shelf and then identify and pick selected objects from the shelf into a box, in order to fulfill an Amazon order.

In 2015 and 2016, Team MIT's competition entry consisted of a robot with a custom end-effector, consisting of two rigid fingers mounted on a force-controlled parallel jaw gripper. Each finger consisted of a plastic body that ended with a compliant, actuated fingertip which allowed the fingers to reorient their positions for different picking strategies, which we call manipulation primitives. Team MIT's end-effector is able to suction objects, grasp objects between its fingers, and scoop items lying on a flat surface. These primitives were, in combination, able to pick up all items in the challenge.

Each primitive relies on vision and sensor data to successfully manipulate objects. For example, grasping uses photogate sensors to detect if an object is grasped and uses force feedback in the parallel jaw gripper to determine when to stop closing the gripper. Scooping relies on strain sensors in the fingertips to detect contact of the back of the shelf. Suction uses hall effect sensors to detect movement of the suction
cup and relies on vacuum flow sensors to detect seal on an object. Manipulation is dependent on numerous methods of tactile sensing; sensor information is used to create a model of grasping, scooping, or suction, and the system needs to be robust enough to determine what primitive should be performed on each object.

1.2 Description of tactile sensing

Tactile sensors are an important aspect of perception, enabling robot hands to manipulate objects and explore the environment. While manipulating objects, tactile sensing can provide contact forces and contact motion between the finger and object, which is useful to understanding slip and friction. These properties can be used to optimize and model the "quality" of a grasp. Tactile sensing is also useful for exploration, where object properties like friction, surface texture, mass, etc can be obtained to identify an object.

Sensors often used in manipulation include tactile array sensors, which provide contact shape and pressure in a discrete grid [3]; joint angle sensors, which determine
position of fingertips; force-torque sensors to aid in providing contact forces and torques; and dynamic sensors that detect motion or vibrations [4] to measure slip or friction. All these sensors are able to provide discrete data, but the challenge is in integrating the data to come up with a concrete grasp model.

Figure 1-2: Uses of touch sensing in manipulation. Each sensor is shown on the left, linked to the sensor data they output. [3]

Challenges in tactile sensors also lie in reducing sensor size to fit all the sensors described above in a small area, while also providing enough resolution for each sensor to reconstruct pressure, contact area, or slip. The gelsight sensor in the current configuration is unique in its ability to provide multiple tactile measurements: surface texture, contact area, shear, slip, and deflection.

1.3 Gelsight

Gelsight is a tactile sensor developed at MIT CSAIL that can capture the surface texture of materials up to submicron accuracy [5]. It consists of a clear gel with a specular paint coating. When an object presses against the side of the gel with the reflective coating, the coating’s skin distorts to conform to the objects surface. The
camera, placed on the other side of the gel, is able to record an image of surface features (fig 1-2). In addition, depth information can be captured at sub-micron resolutions by shining multiple colored lights on the surface and using a photometric stereo algorithm to reconstruct the surface topography. Gelsight has also been used in applications for recovering shear, slip, and object hardness, but these fingertip sensors have been created for non-compliant fingertips [6]. The addition of a compliant finger, as is used for grasping and scooping primitives, makes localizing the manipulated object more difficult.

![Figure 1-3](image)

Figure 1-3: (a) A cookie is pressed against the skin of a gelsight block. (b) The skin is distorted, as shown in this view from beneath. (c) The cookie's shape can be measured using photometric stereo and rendered at a novel viewpoint. [5]

### 1.4 Statement of Purpose

This thesis is focused on the application of integrating gelsight into the gripper for the Amazon Robotics Challenge. With a gelsight sensor, we sense three different measurements. First, we use tracking dots on the gel to determine the deflection and side profile of the compliant fingertips. This information is useful in localizing the object in grasping, scooping, and exploring. Next, we use small dots imprinted on the elastomer to measure local shear. Lastly, we can use the gelsight-developed technique to determine surface texture and contact area.
Chapter 2

Development

We discuss the ways that a robotic fingertip can use a sensor like gelsight, using the Amazon Robotics Challenge (ARC) as a model. First, we describe the gelsight finger design, as it applies to the ARC. Next, we detail the tactile properties that the fingertip gelsight sensor can measure. Then, we discuss the use of each manipulation primitives and how they could use gelsight's sensor. Lastly, we discuss the vision processing algorithms that are used to recover these properties.

2.1 Finger Design

Team MIT's 2016 end-effector is composed of a parallel jaw gripper, on which are mounted two rigid fingers. An actuated fingertip, composed of 2 compliant spring steel pieces in the shape of a triangular prism, lies at the end of each finger. This shape is useful in scooping because it acts like a spatula; the sharp fingertip is able to slide beneath objects, and the increasing ramp on the bottom finger allows the object to slide up the finger body. The fingertip's compliance is used to press down and preload against the shelf, to maintain flushness against a scooping surface. Compliance, however, complicates the integration of most tactile sensors.

A piece of gelsight is placed on the inside side of the top finger (Fig 2), where it can make contact with the manipulated object when grasping. To incorporate gelsight on a compliant spatula, we propose creating the actuated fingertips from transparent
Figure 2-1: Placement of gelsight on ARC end-effector and illustration of fingertip spatulas

polycarbonate. This is so that a camera on the interior of the finger body can see the surface image of the gelsight piece attached to the outside. Maintaining a clean and transparent interface between gelsight and the plastic is necessary for capturing clear images of the object’s surface. The bottom side of the gelsight will be covered in small dots for tracking shear and large markers to track deflection.

2.2 Gelsight use in Manipulation Primitives

The design of the gelsight sensor allows us to recover shear, contact area, deflection, or compliance. Understanding manipulation primitives informs us how we should better use or integrate gelsight. Each section introduces a primitive and then addresses how gelsight could provide sensing.
2.2.1 Grasping

Grasping is used in both horizontal gripping, as in picking a book from a shelf, or vertical picking, as in grabbing a cookie out of a jar. In horizontal grasping, the gripper slides its fingers around an object, actuating the fingernails to point inwards to increase contact area. Then, it uses the parallel jaws to compress the object, with friction preventing the object from slipping out of the grasp. See fig. For vertical picking in a cluttered environment, as in the APC, the fingers point straight and are thin enough to slide between other objects, to encompass the picked piece.

Currently, grasping is sensorized with infrared photogates, which are mounted in the finger body. When an object is grasped, the photogates detect that they are blocked, and we can tell where the object is in the grasp, based on these discrete measurements.

Gelsight can provide additional sensor information to model grasp stability. When starting a grasp, the gripper outfitted with gelsight on the inside part of a finger can preliminarily close onto the object, allowing it to use gelsight contact area and texture data. This can confirm the object’s pose. We can summarize this data into large features like edges or flat surfaces, to confirm the robot’s vision model. In addition, contact area and texture can be used to create a model of grasp stability. For example, if we sense a small contact area in grasping (could be categorized as a corner), and it rests primarily in the front fingertip portion, the grasp is unstable.
Lastly, shear data can be used to determine if the object is falling or slipping within the grasp; this is particularly useful, since grasping relies on friction.

### 2.2.2 Scooping

![Diagram of scooping primitive](image)

Figure 2-3: Side view of scooping primitive. The gripper preloads the fingernail to slide it underneath the object. Then, it slides forward until the object pushes against the back wall and slides higher onto the finger body. [2]

Scooping is used to pick up objects that are lying flat on a surface. In scooping, the gripper first preloads the fingernail against the bottom of the shelf, in order to create a ramp from the bottom of the shelf up to the finger body. Then, it slides the fingernail underneath the object and pushes it against the back wall, to push the object onto the more stable, flat finger body. The surface of the gripper is covered with gun tape, which allows a higher static coefficient of friction, useful in grasping, but a low kinetic coefficient of friction, allowing objects to easily slide up the finger.

Scooping is currently sensorized with a strain gauge on the compliant fingertip to detect contact when it hits the back wall. This would trigger the gripper to close and then withdraw from the shelf, and this prevents the gripper from crashing into the
back wall.

Gelsight in scooping could provide sensor information to locate the object while scooping, utilizing contact area, shear, and deflection. Contact area sensing from gelsight could tell us the object's location on the fingernail. This is crucial to holding the object in a stable grasp; if the object is not far enough up the spatula's ramp, it risks falling out of the grasp. Shear sensor data can indicate the object's motion up the fingernail, providing velocity data, as well as slip or stick behavior. Lastly, gelsight's deflection data can replace the current strain gauge to measure scooping preload, and help the robot detect the bottom of the shelf.

2.2.3 Suction

![Figure 2-4: Side view of suction primitive](image)

Suction is used to pick up objects that may not be easily grasped or scooped. In suction, our suction cup is mounted near the end of the gripper. It is able to suction downwards, useful in picking from a shelf, and suction forward, useful in vertically picking from a cluttered bin. In addition, the suction system should be able to stow the cup, so that it does not stick out of the gripper. In past APC designs, the suction cup used hall effect sensors to check if the suction was stored, and whether or not an
object was under suction.

Gelsight in suction would only be useful at the suction cup, since gelsight’s strengths lie in contact sensing. If the suction cup bellow was made from gelsight material, and a camera could be positioned behind, contact area sensing could be useful to detect contact to understand whether an object is picked up, and shear could be used to determine if the object is slipping in the suction’s grasp, and if needed, suction force could be increased.

2.2.4 Exploration

Lastly, gelsight as a contact sensor would be useful in exploring the surrounding environment. If mounted on a gripper finger, it could be used to probe the environment to get object pose and location through deflection and contact area. Locating edges and flat areas would be useful in estimating a grasp pose. Conversely, if the object’s position was known, the robot could use this contact as localization. Previous manifestations of APC grippers have only explored by determining an environment’s boundaries. Strain sensors have prevented the gripper from crashing into walls.
Chapter 3

Proposed Approach

3.1 Manufacture of fingertips

To create the fingertips, 1 mm transparent polycarbonate is bent and attached to the finger body. Illumination and camera position are critical to recovering an image with enough contrast to decipher image details. The thin, triangular fingertip shape made it tricky to recover the camera image, as the camera must be located in a static position. We use a small, glass mirror to enable the camera, located in the finger body, to view the gelsight image.

![Figure 3-1: The camera and mirror placement to recover gelsight image](image)

To keep lighting conditions consistent, we glue an elastic cloth to the sides of the plastic fingertips to block out all outside light. Then, we can light the interior with ultra-bright white LEDs. To counter the problem of light directly reflecting off
the shiny plastic, we sand the top side of the finger to create a rough surface where light can scatter. In addition, we paint the insides white to reflect any light, and we position the LEDs far behind the camera so we do not produce direct reflections. The lighting conditions are finely tuned to capture the multiple markers that are used to recover shear, contact area, and deflection.

3.2 Manufacture of gelsight

The gelsight pad is composed of silicone rubber XP-565 from Silicones, Inc. in a 15:1 ratio, molded to be flat where it adheres to the spatula, and domed on the bottom side for point contact with objects. Gelsight is adhered to the bottom face of the fingertip with silicon tape, being careful to create a clear bond that doesn’t decrease the quality of the camera image seen through the polycarbonate-gel interface. Then, a thin coating of Silbione LSR 4330, a liquid silicone rubber with 30 durometer, is applied on the surface to increase durability as the sensor touches multiple objects.

Our tactile sensor can recover 4 properties: contact area, shear, deflection, and compliance.

To recover contact area data, we capture images of the gelsight’s surface. Gold paint is brushed on the domed, bottom side of the gelsight that comes in contact with objects. This specular coating on the membrane shows perturbations on the membrane surface. Because the membrane takes on the object’s 3D topography, the camera placed behind the finger will be able to infer the texture of the object being probed, as well as the pressure distribution. Objects that are pressed harder look darker than objects that are lightly touched.

To track shear deformation, a grid of small 1 mm dots were placed on top of the specular paint with water transfer paper. These dots move with the layer of specular paint, and we can capture images of the grid, which create a displacement field. We compare the movements of these dots over time, to recover local movement on the object, which can be used to categorize slip.

To track deflection of the fingertips, we place six large markers between the trans-
Figure 3-2: We show the gelsight finger covered in specular gold paint and a grid of small dots for tracking shear. An edge contact area feature can be seen.

Figure 3-3: We use 6 large checkerboard markers to recover fingertip deflection.

parent plastic fingertip and the gel sample. This is distinguishable from the higher frequency small grid of dots by their shape. As the spatula bends, we can use these six dots to recover the side profile of the right and left side. We can reconstruct the pose of these dots with computer vision techniques, by creating a mapping between the ground truth and dot location and size. Knowing the spatula's deflection is useful in locating the manipulated object in space.

Lastly, we can also recover an object's compliance. Knowing the fingertip's elastic modulus and measuring the force applied during contact with the parallel jaw gripper's force-torque sensor, we can calculate the compliance of the object at that point.

3.3 Electrical Setup

The camera inside the finger body is a raspberry pi spy cam, which possesses a wide depth of field, because of the small pinhole lens. In addition, it is able to capture
Figure 3-4: Electrical system to communicate between raspberry pi and computer up to 1080p images at 30 fps, and easily interfaces with the raspberry pi, the microprocessor that also controls motors and other sensors on the gripper. Simple image processing to calculate deflection and track shear dots is preliminarily done on the raspberry pi, and then this information is communicated to the main robot computer through ROS.
Chapter 4

Vision processing to recover deflection data

In order to recover the deflection of the spatula, we use a group of distinct markers, placed beneath the gelsight pad. We can create a mapping between the markers' image coordinates and their real coordinates. These real coordinates can then be used to model a side profile of the fingertips. Bending is taken into account, as there are dots spanning the width of the spatula, and thus, a unique side profile for each side of the spatula.

We use as ground truth a Vicon motion capture system, which captures very accurately the pose of small reflective markers. Since the vicon dots are beneath the gelsight, they have a low chance of being perturbed by contact. In addition, the camera position is held constant, and the lighting configuration is made consistent. By creating a mapping between the vicon real coordinates to image pixel locations, we can later reverse this mapping and derive the real coordinates from image pixel locations. We evaluate the accuracy of such a mapping on an example deflected spatula. Although this spatula is larger than the real fingertip and holds more deflection configurations, it is a test case for creating a real mapping with the ARC fingertips.
4.1 Experimental Setup

The experimental setup consists of a fixture with a simple web camera, the polycarbonate spatula, the vicon dots. During the experiment, the camera images and the vicon 3D are taken simultaneously in ROS, to synchronize the time-stamps. In addition, because of the reflection of the polycarbonate, we take the data in low light.

![Figure 4-1: On the left, we can see the overall setup used to create the mapping. On the right, we see a closeup of the vicon markers](image)

The mapping is created by collecting data on the gelsight spatula setup while deforming the spatula to many different configurations, while gathering image and vicon data through ROS. Then, conventional image processing toolboxes in MATLAB are used to process each images, thresholding images based on color, and applying three erosion and dilations to eliminate small features and produce a cleaner binary image of the tracking dots on the spatula. A circular Hough transform is used to identify all circle centers. These circle centers are denoted in pixel coordinates, which are then mapped to ground truth values from Vicon.

We collected a data set of 1290 images over 43 seconds. 1201 of those images were able to get a good centroid estimation. To verify the mapping, we divided this data set into 5 parts. We processed the mapping on \( \frac{4}{5} \) of the data set and then calculated the accuracy of that mapping on \( \frac{1}{5} \) of that set. To perform the mapping, a nearest neighbor search is done to correlate those 10 pixel coordinates with the closest array of pixel coordinates in the data set. To calculate accuracy, we count the number of
Figure 4-2: On the left is the original image, with centroid points overlaid in red. On the right is the image after thresholding.

images where all points are within a distance threshold. This mapping is performed five times, for each part, and the accuracy is averaged to be 75.6%.

<table>
<thead>
<tr>
<th>Group 1</th>
<th>Group 2</th>
<th>Group 3</th>
<th>Group 4</th>
<th>Group 5</th>
</tr>
</thead>
<tbody>
<tr>
<td>93.7 %</td>
<td>65.8 %</td>
<td>68.2 %</td>
<td>54.9 %</td>
<td>89.7 %</td>
</tr>
</tbody>
</table>

Table 4.1: Accuracy of each group. The average accuracy is 74.5%.

This accuracy is insufficient for use in manipulation primitives, as having a model of the spatula’s deflection is essential to understanding shear or contact area. Having an accurate deflection model only 75% of the time is not accurate enough to localize the fingertip. However, there are many ways that this method can be improved. The dip in accuracy in groups 2, 3, and 4 show that the data is perhaps structured or that much of the mapping is collected from that subset. This means that collecting more data would be helpful. We plan also to investigate other methods to learn the mapping from image pixels to curvature directly, hopefully yielding better results, benefiting from the overall structure of the image, rather than just local deformations.

We plan to further this by performing the same analysis on the real fingertip shape. This means moving from the usb camera to a raspberry pi setup, working
with fewer vicon dots, and collecting more data to create a lookup table.

Figure 4-3: Deflection mapping setup on a more accurate fingertip shape
Chapter 5

Conclusion

In this thesis, we were able to evaluate the use of a gelsight sensor for manipulation primitives, design for its use in Team MIT’s gripper platform, and create a testing platform to evaluate the accuracy in reconstructing deflection data.

Reconstructing deflection is the first step to localizing the object. Further work will be done on reconstructing contact area and shear from the raspberry pi images, primarily using edge detection and spatial gradients to label the object with classifiers.

These classifiers can be used in the three manipulation primitives of grasping, scooping, and suction, to provide closed-loop feedback on the timing of each primitive, as well as create a model of when each primitive should be used. Gelsight’s use as a manipulation sensor is able to provide contact area, texture, shear, and deflection in a compact area.
Bibliography


