Touching is Believing: Sensing and Analyzing Touch Information with GelSight

by

Rui Li

Submitted to the Department of Electrical Engineering and Computer Science

in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

June 2015

Signature redacted

Author .....

Department of Electrical Engineering and Computer Science

May 20, 2015

Signature redacted

Certified by .....

Edward Adelson

Professor

Thesis Supervisor

Signature redacted

Accepted by .....

Leslie A. Kolodziejski

Chairman, Department Committee on Graduate Theses

Touching is Believing: Sensing and Analyzing Touch Information with GelSight

by

Rui Li

Submitted to the Department of Electrical Engineering and Computer Science
on May 20, 2015, in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy

Abstract

For robots to perform advanced manipulation in a world of unknowns, touch is a critical source of information, and a high-quality tactile sensor is essential. However, existing tactile sensors generally are low-resolution and/or non-compliant, making it difficult to capture detailed contact information for manipulation that humans are very good at.

GelSight was first developed a few years ago to capture micro-scale surface topography and converts pressure patterns to images, making it promising for high-quality tactile sensing. However, the original devices were big, relatively slow, and expensive for robotic applications. In this work, we developed a new tactile sensor based on GelSight, which we call fingertip GelSight sensor, that is much more compact, faster and less expensive. Despite that, the fingertip sensor has high resolution, on the order of tens of microns, high compliance and high sensitivity. We demonstrated its unparalleled capabilities as a new-generation robotic fingertip for manipulation, in terms of localization and manipulation of small parts, normal and shear force estimation, and slip detection, as well as for material recognition, in terms of 3D surface texture classification. With image processing and machine learning techniques applied on the tactile images obtained, the fingertip GelSight sensor opens many possibilities for robotic manipulation that would otherwise be difficult to perform.

Thesis Supervisor: Edward Adelson

Title: Professor

Acknowledgements

I would like to express my sincere gratitude to my thesis supervisor, Prof. Edward Adelson. Words cannot express how grateful I am for all his trust, advice and learning opportunities he has provided in the past 4 years. He is a thoughtful friend, an inspiring teacher and a great researcher that I have learnt so much from. His critiques helped me think more thoroughly about the problems to solve, his guidance helped me with how to do research, and his tremendous support and understanding helped me go through the most difficult times in my life.

I would like to thank my thesis committee, Prof. William Freeman and Tomas Lozano-Perez, for their valuable feedback and insightful questions. Their help made an important part of my PhD journey.

Throughout the development of the various projects, I feel very fortunate to have worked with a group of talented and insightful collaborators who have provided great help in various aspects that made my research experience a lot smoother. Mandayam Srinivasan, a leading expert in haptics from MIT Touch Lab, has provided lots of insights in many of the projects; Prof. Robot Platt from Northeastern University is an expert in robotics and has been extremely helpful and insightful in pushing GelSight's applications for small parts manipulations and beyond. I still remember when Robert first approached us to explore GelSight's capabilities in cable insertions, which was proven to be a huge success in demonstrating GelSight's potential. Great thanks also go to my group mates Wenzhen Yuan and Xiaodan Jia who worked with me on some of the GelSight projects, helped to make gels, set up experiments and so on. Their help is highly appreciated and made my research experience smoother and more enjoyable. I would also like to thank Nathan Roscup, Andreas ten Pas, and Mordechai Rynderman who helped with 3D printing of the GelSight housing, or control of Baxter robot for the USB insertion tasks.

Great thanks also go to Andrea Antonello for assisting some experiments for force estimations and 3D CAD rendering of the fingertip GelSight sensor.

I also want to thank my other group mates, Phillip Isola, Shaiyan Keshvari, Derya Akkaynak, and Bei Xiao, who made my experience so much delightful. Our lab assistant John Canfield provided great assistance by ordering items needed for my research and so on. He was always there and ready to provide help.

I feel very fortunate to have met many great friends throughout the MIT journey: Lei Dai, Yinqing Li, Yu Xin, Hang Chen, and so on. You guys made my life filled of fun, laught and unforgettable memories. MIT is a wonderful place with amazing people, and the 6-year experience here is an absolutely precious part of my life.

Last but not least, I want to thank my family for their selfless love and support that guided me through the many years that passed and more years to come. Behind every step of my journey, I have them with me and I am full of energy.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Imagine one day in the future when robots live and work with humans. At 7:30am, your robot assistant Emma wakes you up, and asks what you want for breakfast. While you wash up, she then starts to get busy in the kitchen ***to play with all sorts of cookers*** *to prepare your favorite breakfast set: scrambled eggs, bread, salad and coffee. Before you go out to work, Emma tells you that today is cold, and suggests you wear something thick. You tell her to* ***bring the gray, soft, and thick sweater****. She brings it over, helps you wear it up, and then* ***opens the door*** *for you. When you arrive at your work place, in the production lines, many other robots like Emma are already lining up to be ready for work. They do* ***cable insertion*** *with high efficiency to check the electronic devices work correctly. At 4:30pm, you get off from work, and enjoy the milk tea Emma has already prepared for you. For dinner, Emma* ***makes very delicious dumplings****, by hand.*

The above scenario may be a very small part of what robots can do one day, with touch capability as an important function and a critical source of information. The field of robotics has been expanding from production lines to include more complex

environments with lots of unknowns such as homes, offices and hospitals. For robots to function well in a world of unknowns and reason what to do when interacting with humans and wide range of tools, such as cookers, in the real-world environments, versatile autonomous intelligent robots are required. To interact with humans more naturally and perform human-like manipulation tasks, advanced manipulation capabilities are required, in order to manipulate objects while actively sensing and reasoning about the environment. To do so, it is important that robots have an interface to provide forces and positions about the contact information between robots and the objects they are interacting with. A high-quality tactile sensor becomes very essential.

In this work, we aimed to build tactile sensors that are similar to human fingertips with the following properties: high sensitivity to light touch, high resolution, and compliant. In addition it is desirable to have information about force and torque.

We propose to design a new kind of tactile sensor based on GelSight [1-2], which has extremely high resolution and sensitivity, and can mimic human touch in many ways, such as sensing forces and torques, detecting slips, and obtaining surface textures. With our new compact design, we are able to install a GelSight sensor onto a Baxter robot as a fingertip, and demonstrate its great performances for different types of applications such as small parts manipulations and material perception. With image processing and machine learning on the tactile images obtained by GelSight, it opens many possibilities for robotic manipulation based on tactile sensing [3].

1.1 Background

Humans interact with the world around them using five senses: sight, hearing, taste, smell and touch. Among them touch is an important source of information. We use our hands to touch and grasp objects for manipulations, to touch and feel the materials of the clothes, to press keyboards of computers and smartphones, etc. We use touch so often that we almost take this capability for granted. However, for robots and other machines, developing the sense of touch has been very challenging.

1.1.1 Human Touch

Before we go into the robotic tactile sensor, it is necessary to understand how human touch works. For humans, the tactile sensory signals due to contact are given by mechanoreceptive afferent neurons (mechanoreceptors) innervating the outer layers of the skin [21]. Four different types of afferents have been identified, with different functions and sensing ranges. Two types of fast adapting afferents (type I and type II) respond to temporal changes in skin deformations (dynamic) while the other two types of slow adapting afferents respond to sustained deformations over time (static). The type I afferents are located in the dermal-epidermal boundary, with small and well-defined receptive fields and highest density at the fingertips. The type II afferents are in deeper layers of the skin, with larger and more diffuse receptive fields, and are throughout the fingers and palms of the hands [21] (Fig. 1-1).

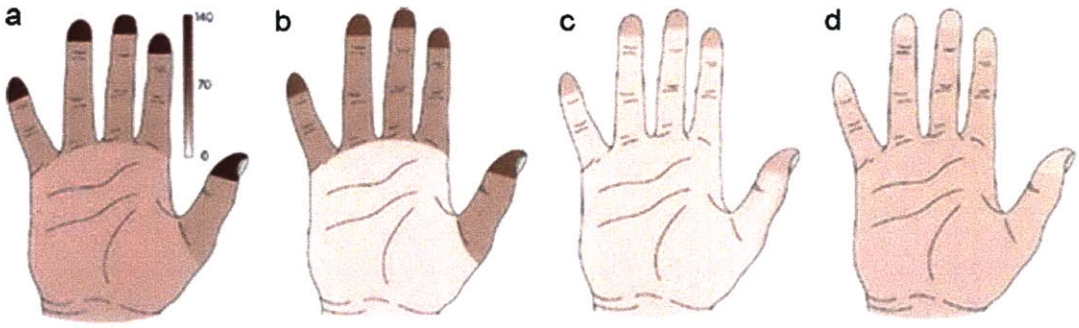


Figure 1-1 [21]: Density of mechanoreceptors (afferents per cm^2) in the hand. (a) Fast adapting type I, (b) slow adapting type I, (c) fast adapting type II and (d) slow adapting type II. Color coding for all four figures is shown in (a).

1.1.2 Robotic Tactile Sensors

In the past few decades, many tactile sensors for robots have been developed with the goal of mimicking human fingertips, with different advantages and disadvantages, including resistive sensors, capacitive resistors, piezoelectric sensors, optical sensors, and organic field-effect transistors (OFETs). [23] provides a review of various tactile sensors. Resistive sensors are developed with the principles of changing resistances with changing forces, and are the predominant choice for tactile sensing. Examples include micromachined strain gauges [24-33], micromachined piezoresistors [34-40], conductive

polymers and fabrics [41-49], conductive elastomer composites [50-55], and conductive fluids [56, 57]. In addition, several publications can also be found on piezoelectric and capacitive sensing techniques, as well as on different optical systems.

However, prior sensors fall short in various ways. They generally have low spatial resolution and/or non-compliance. Low spatial resolution (e.g., 1x1 to up to 32x32 with millimeter spatial resolution) makes it difficult to capture detailed force distribution for manipulation and surface textures for material perception. Non-compliance in some cases makes it difficult to conform to the object's surface for stable grasps. To solve some of the challenging problems in tactile sensing, some new tactile sensors are needed.

1.2 Contributions

The special properties of the GelSight sensor make it very promising to be used as a robotic fingertip, but the original GelSight devices were so large that they were not really designed for robotic applications. In this work, we describe a new kind of tactile sensor based on GelSight, called fingertip GelSight sensor, that has high resolution and high sensitivity, is fast and inexpensive, and can be made very compact. This thesis focuses on several research questions that aim to solve some of the key challenges in tactile sensing for robotic manipulation and material perception. In this thesis, we describe the following contributions:

- **Development of a robotic fingertip based on GelSight**
  - Redesigning the GelSight device to make it very compact for use on a Baxter robot gripper while keeping its high resolution and compliance.
  - Reducing the overall cost of such a sensor by a factor of 10-100 times.
  - Writing APIs that readily give GelSight outputs including the raw image data, surface normal, 3D height map, force field, and slip detection results.

- **Robotic fine manipulations with GelSight tactile sensing**
  - Developing algorithms and programming for localization and manipulation of small parts using a fingertip GelSight sensor.

- Demonstrating fine manipulations for a USB insertion task using a fingertip GelSight sensor equipped on a Baxter robot.

- **Force estimations and slip detection**
  - Using height map to infer normal force qualitatively, as it is a direct reflection of the amount of gel deformation.
  - Developing algorithms and programming to infer shear force by tracking movements of the markers on the GelSight membrane.
  - Developing algorithms and programming to do slip detection based on the normal and shear force estimated.
  - Programming the Baxter robot equipped with a fingertip GelSight sensor to perform tasks that utilize the force and slip information, such as picking up objects of different properties, ranging from bottles, eggs, and potato chips to leaves and cups with various amounts of water.

- **Material perception with GelSight tactile sensing**
  - Developing algorithms to classify 3D surface textures obtained from GelSight, and achieved a high success rate.
  - Programming for fast and accurate material recognition.

1.3 Thesis Outline

Chapter 1 describes the background and contributions of this work. Chapter 2 first gives an overview of the GelSight technology, and then describes details of the new fingertip GelSight sensor developed, including its hardware, software and performances. Chapter 3 discusses an application of the fingertip GelSight sensor in robotic fine manipulations of small parts by demonstrating a USB insertion task with a high success rate. Chapter 4 presents methods for estimating the normal and shear forces as well as detecting slip for a fingertip GelSight sensor, and for using such information in robotic grasping tasks. Chapter 5 shows algorithms and experiments for classification of surface textures using a portable GelSight device previously developed, and the results can be readily applied to a

fingertip GelSight sensor for material recognition. Chapter 6 concludes the thesis and suggests future work for improvements.

Chapter 2

Fingertip GelSight Sensor

This chapter gives an overview of GelSight, and the development of the fingertip GelSight sensor for robotic applications.

2.1 GelSight Overview

GelSight is a novel tactile sensor that converts information about surface shape and pressure into images, which was first proposed by Micah Johnson and Edward Adelson in 2009 [1], and later refined in 2011 [2]. In its simplest form, GelSight is a piece of clear elastomer coated with a reflective membrane, along with a camera and illumination. When an object is pressed against the membrane, the membrane deforms to take the shape of the object's surface, which is then recorded by a camera under illumination from LEDs located in different directions (Figure 2-1). A 3-dimensional (3D) height map of the surface can then be reconstructed with the shape-from-X method [1]. Figure 2-2 illustrates the basic GelSight principle.

In [1], the illumination system consisted of red, green, and blue lights at different positions that can provide three shaded images. The surface can be reconstructed using a photometric stereo algorithm tailored to the sensor. In [2], the reconstruction was improved with a combination of improved sensor materials, illumination design, and reconstruction algorithms. Also, two configurations of the GelSight devices were presented: the bench configuration with ultra high resolution (Figure 2-4(a)), and the portable configuration with relatively low resolution but fast reconstruction in seconds (Figure 2-4(d)). For robotic applications, however, both configurations are still too large and slow as robotic fingertips. We propose a finger configuration that is much more compact and fast in Section 2.2 below (Figure 2-3(c)).

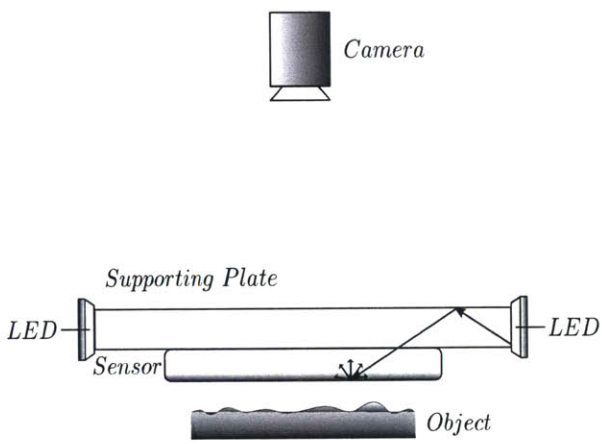
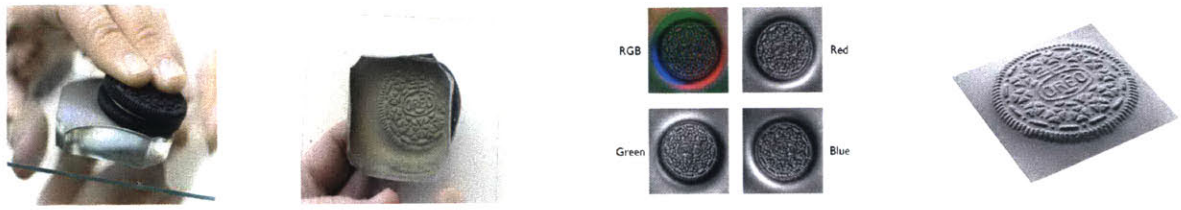


Figure 2-1: GelSight schematic. A typical GelSight system consists of an elastomer coated with a membrane, the deformation of which is captured by a camera under illuminations by LEDs located in different directions.

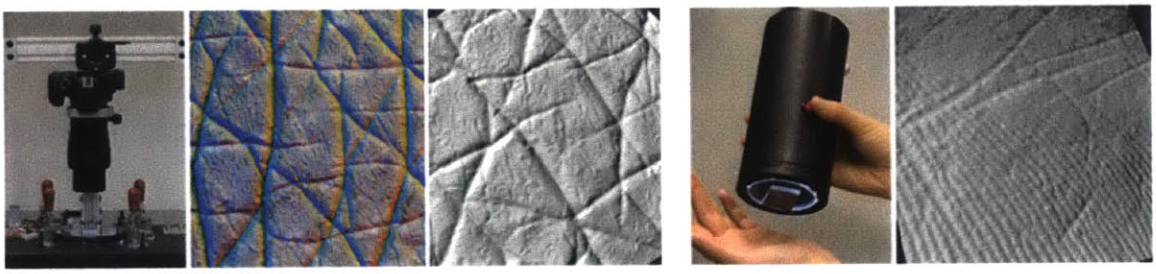


(a) (b) (c) (d)

Figure 2-2 [1]: GelSight illustration. (a) A cookie is pressed against the membrane of an elastomer block. (b) The membrane is deformed to the shape of the cookie surface. (c) The shape of the cookie surface is measured using photometric stereo and rendered at a novel viewpoint.

(a) (b) (c)

Figure 2-3 [1]: Different configurations of GelSight devices. (a) is the box configuration, (b) the portable configuration, and (c) the fingertip configuration.



(a) (b) (c) (d) (e)

Figure 2-4 [2]: The microgeometry capture system consists of an elastomeric sensor and a high-magnification camera (a). The retrographic sensor replaces the BRDF of the subject with its own (b), allowing microscopic geometry (in this case, human skin) to be accurately captured (c). The same principles can be applied to a portable system (d) that can measure surface detail rapidly and easily; again human skin (e).

The GelSight sensor has many important properties that make it attractive to be used as a new-generation tactile sensor. First, the sensor can give spatial resolution as small as 2 microns, which is limited by the size of the membrane particles used to make the gel. This allows capture of high-resolution surface textures for material recognition [2] and roughness measurement. Second, the sensor is compliant and deformable, and can be used to measure the hardness of a touched surface, ranging from soft surfaces such as human fingers to hard surfaces such as fabric and woods. It was also previously shown that GelSight can be useful in lump detection [4]. Third, the sensor output is not affected by the optical characteristics of the materials being measured as the sensor membrane supplies its own bidirectional reflectance distribution function (BRDF). This allows us to

capture a wide range of material surfaces no matter whether they are matte, glossy, or transparent [2]. Last but not least, GelSight may also be used to measure the pressure distribution across the contact region, as well as shear and slip between the sensor and object in contact. All these properties make GelSight a very promising candidate as a new-generation tactile sensor.

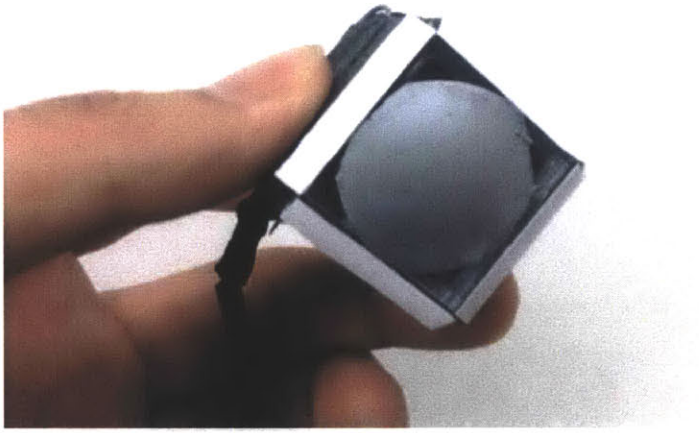
2.2 Fingertip GelSight Sensor

In order to use GelSight in robot manipulation tasks, the sensor must be mounted in the fingertip of a robot hand. However, since prior GelSight designs are physically too large for this and not running in real time, a new design was needed.

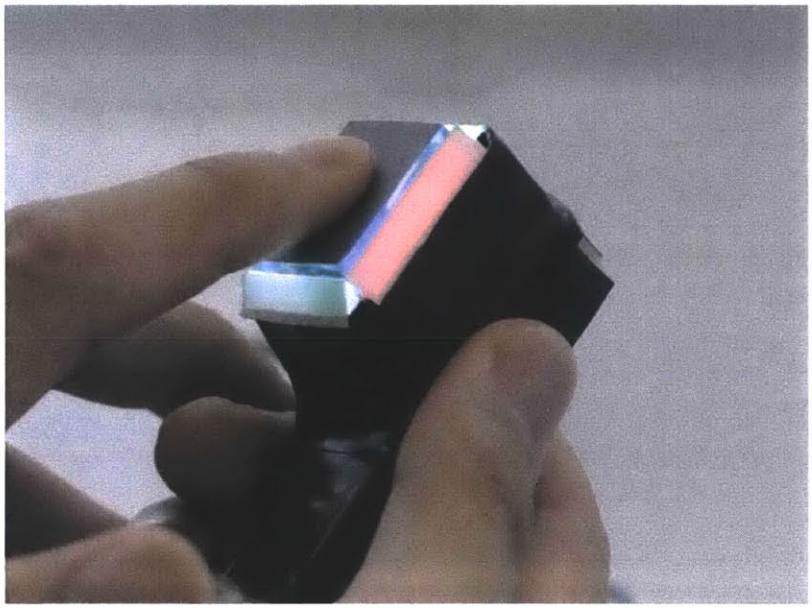
2.2.1 Hardware Design

In our work, we designed a new sensor with much smaller size, comparable to the size of human fingertips, as illustrated in Figure 2-5(a) and (b). The elastomer gel is shown at the bottom of Figure 2-5(c). Above it is a hollow box. At the top of the box, a camera points downward through the box and onto the gel. The camera captures deformations in the gel caused by contact. A key aspect of the design is the way the gel is illuminated. We illuminate the gel from four sides simultaneously in four different colors: red, green, blue, and white. As a result, images of the gel with illuminations from the four different directions can be captured nearly independently in the camera RGB channels. By viewing the contacting membrane from four different directions, it is possible to calculate surface normal on the contacted surface. Then, the height map is calculated using Poisson integration. Since it is important to perform this calculation in real time, we store the correspondence between colors and surface normals in a lookup table. We can calculate the height map at 10+ frames per second in Matlab 2013b on a 2.8 GHz Intel Core i7 running 32-bit Windows 7. Light from each of the four LEDs is directed by light guiding plates into the supporting plate and the gel. As a result, the path length of the light is maximized so as to simulate a more parallel illumination system as assumed in photometric stereo. This improves the accuracy of the resulting height map of the contacted object surface. All together, this sensor is a cube approximately 3 cm on a side

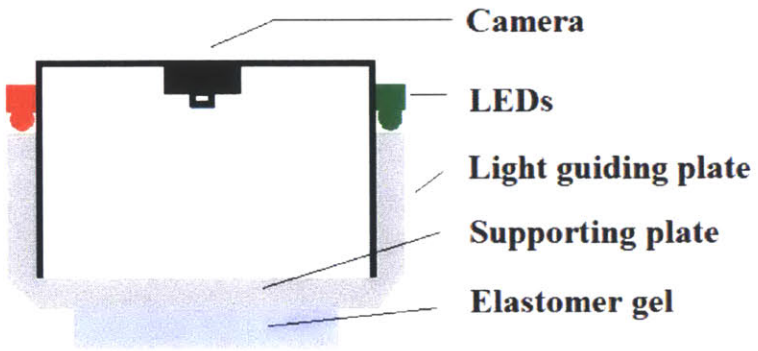
(the tail of Logitech C310 is not shown in Figure 2-5(a)). Smaller and faster versions of the GelSight sensors are possible, and we will further improve the hardware and software to make the sensors more compact, robust, and faster, and give more accurate results.



(a) Fingertip GelSight sensor version 1 that is stand-alone



(b) Fingertip GelSight sensor version 2 that can be mounted on a Baxter robot



(c) Design schematics of fingertip GelSight sensor

Figure 2-5: The fingertip GelSight sensor and its design schematics.

There are some important design considerations for the fingertip GelSight sensor. First of all, the webcam used needs to have controllable parameters for the exposure, white balance and gain control. This is critical for the 3D reconstruction using lookup tables, as we want the color tone of the image to be consistent across different scenarios. Also, we want the size of the webcam to be relatively small. Among the different webcams examined, Logitech C310 satisfies all the above and is used in our system. Nevertheless, a smaller camera that satisfies the criteria can definitely be used to further reduce the system size. Also, a wide-angle lens can make it even more compact.

Figure 2-6: Logitech C310 with case on (left) and with case removed (right).

Second, we used LED arrays of multiple colors, red, green, blue and white (RGBW), to simulate parallel illuminations of distant light sources, as required by photometric stereo. The LEDs are of size 1.8mm with 30-degrees viewing angel, and can be ordered from [58].

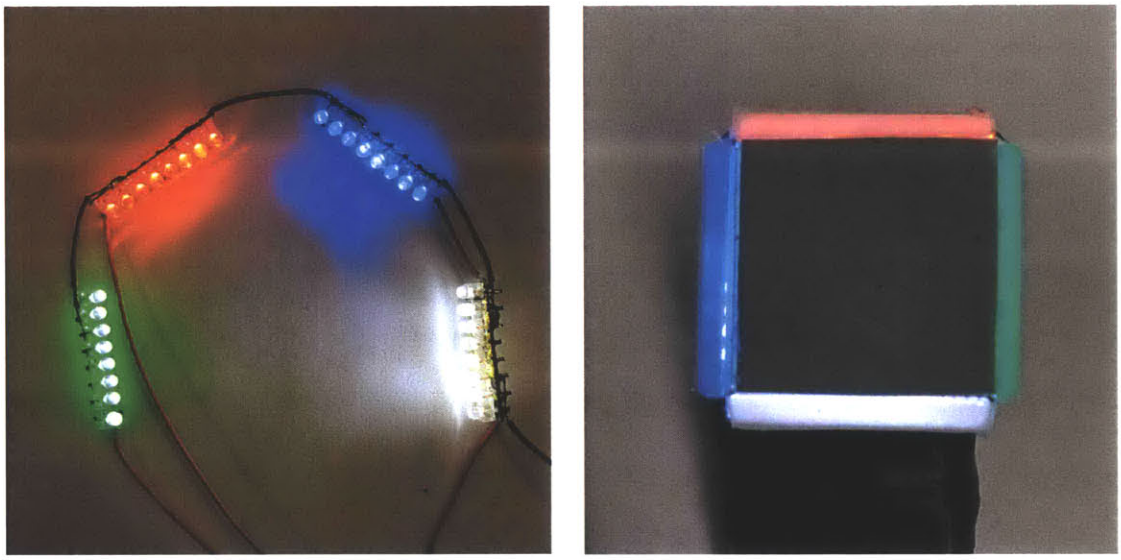
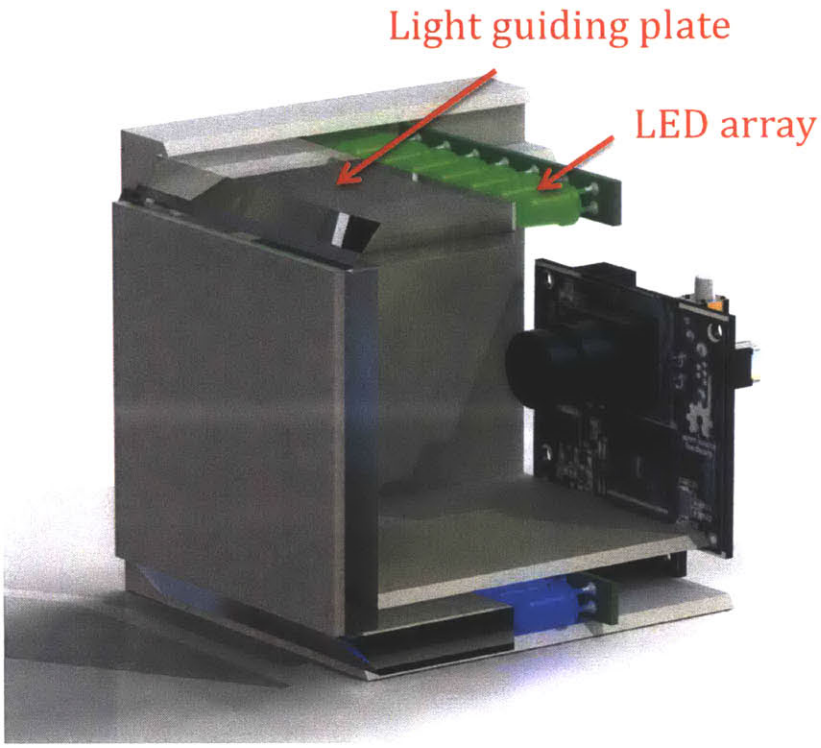
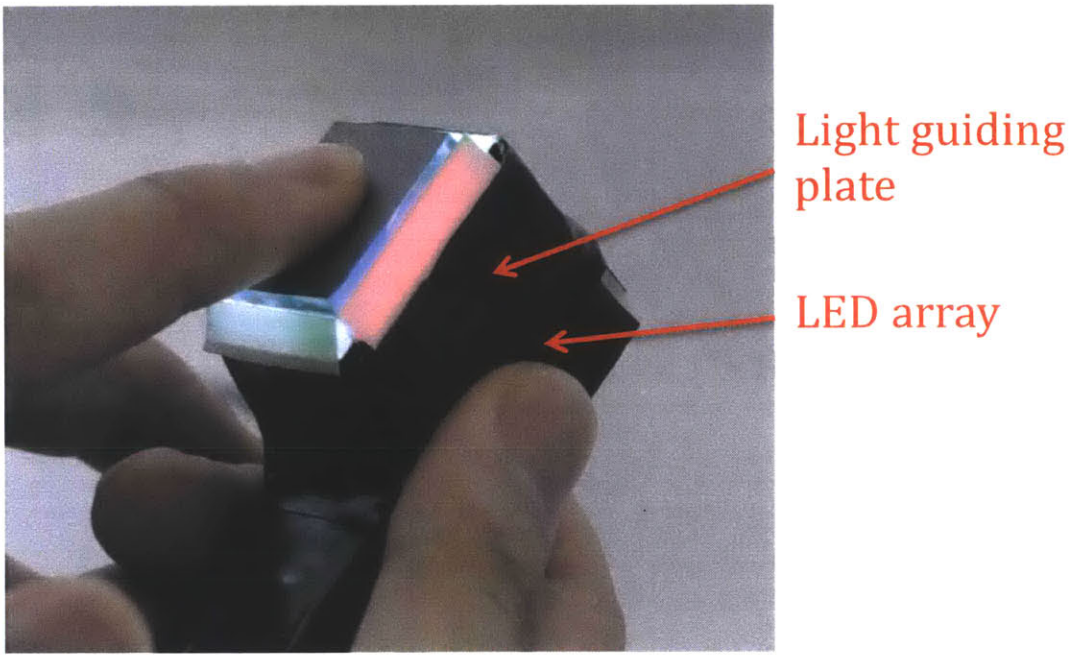


Figure 2-7: LED arrays and their look from the bottom of the fingertip GelSight sensor.

Third, the light guiding plates are transparent acrylics with 1/8-inch thickness placed on each of the 3D printed housing walls, with 45-degree angles on one end to guide and reflect the LED illuminations onto the gels. They are used to increase the light paths while keeping the overall shape of the sensor compact. Figure 2-8 shows the positions of the light guiding plates and LED arrays.



(a) Internal view of the fingertip GelSight sensor



(b) External view of the fingertip GelSight sensor

Figure 2-8: Internal and external views of the fingertip GelSight sensor with positions of light guiding plates and LED arrays shown.

Fourth, we used dome-shaped sensors as an extension of the flat sensors. This allows the robot to deal with even relatively flat surfaces, which is important in robotic applications. This was not considered in previous GelSight systems. Figure 2-9 shows a dome-shaped gel and the corresponding 3D image when pressed against a flat surface.

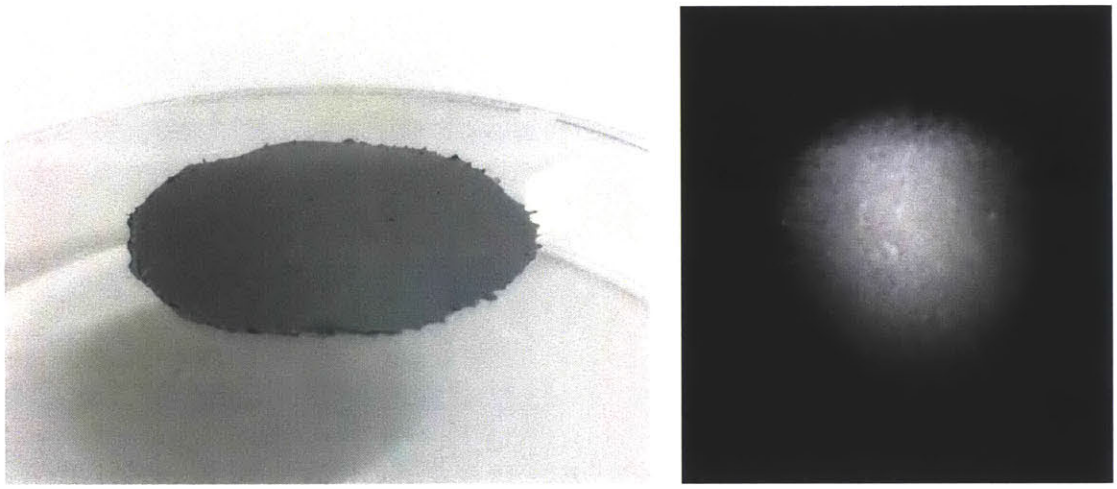


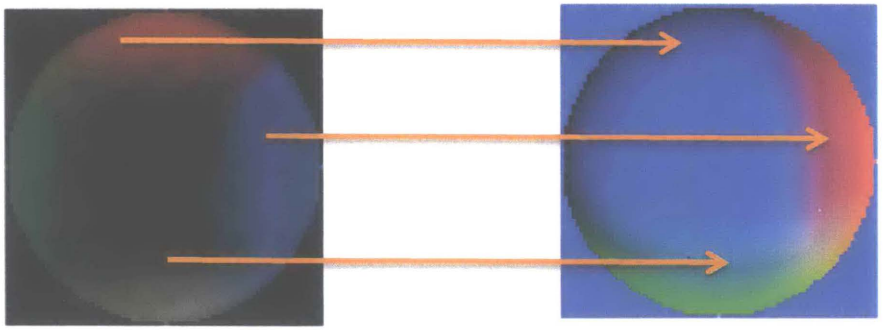
Figure 2-9: A dome-shaped gel (left) and the corresponding 3D image (right) when pressed against a flat surface.

2.2.2 The Algorithm

The algorithm is as follows:

Step 0: Calibration

- Under RGBW illuminations, a pixel has color triplet (R,G,B)
- A semisphere has known surface normals (Nx,Ny,Nz). We can use it as a calibration target
- Subtract background images, and create a lookup table to match (R,G,B) with (Nx,Ny,Nz)



(a) RGB image of semisphere

(b) Corresponding surface normal (Nx,Ny,Nz)

Figure 2-10: The RGB image of a semisphere calibration target and the corresponding surface normal to match to.

Step 1: Create surface normal map (Nx,Ny,Nz)

- Subtract background images from original RGB images
- Convert to (Nx,Ny,Nz), with the lookup table created in Step 0

Step 2: Create gradient map (Gx,Gy)

- Gx = Nx/Nz, Gy = Ny/Nz, when Nz != 0
- Set up a upper limit to (Gx,Gy) when Nz = 0

Step 3: Create 3D height map from (Gx,Gy)

- Solve Poisson equation using Discrete Sine Transform (DST) or Fast Fourier Transform (FFT) [59]

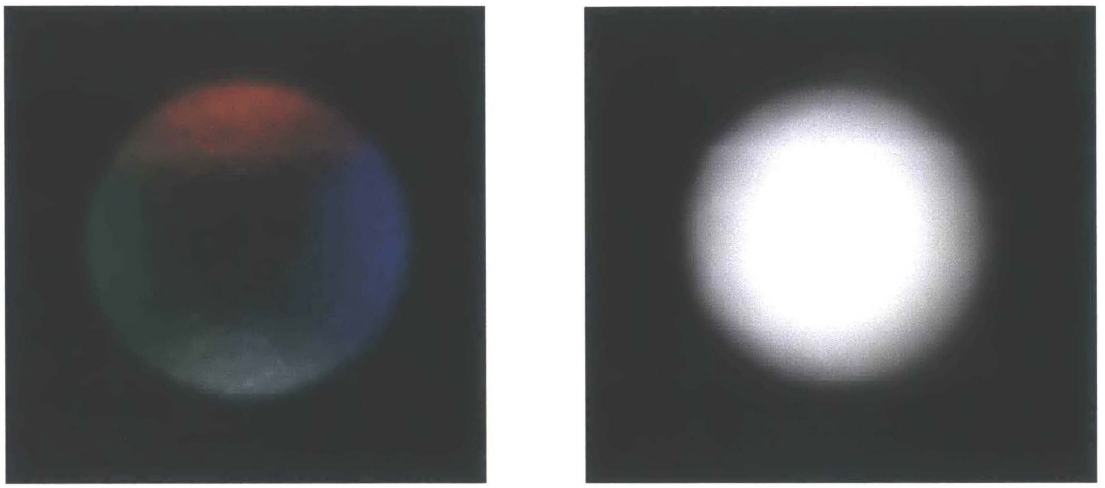
Note that for the size of the calibration balls, it is mostly limited by two factors:

- Thickness of the gel, which limits the maximum diameter so that when pressed onto the gels the whole semisphere can be imaged for calibration, and
- Resolution of the RGB image, which limits the minimum diameter. We want a larger diameter for better quantization of the lookup table for the matching between the RGB triplets and the surface normals because a larger semisphere has more pixels.

By that standard we should use a semisphere with diameter around 2mm. For a larger semisphere for better quantization, we can still do the calibration but need to adjust the

surface normals accordingly depending on the depth of the calibration ball penetrated into the gel.

With the above steps, we have the following RGB image of a calibration semisphere (Figure 2-11(a)) and the corresponding 3D reconstruction or height map (Figure 2-11(b)):



(a) RGB image of the calibration target (b) 3D image of the calibration target

Figure 2-11: RGB image and the corresponding 3D reconstruction for a calibration target that is a semisphere.

2.2.3 Performance

Here are properties of the new fingertip GelSight sensor:

- **Compact**
    - The central part of the fingertip GelSight sensor is now reduced to a cubic box about 30mm on each side, smaller than the previous portable configuration by a factor of about 5 and the bench configuration by a factor of over 50. Figure 2-12 shows the sensor mounted on a gripper of a Baxter robot sold by Rethink Robotics.

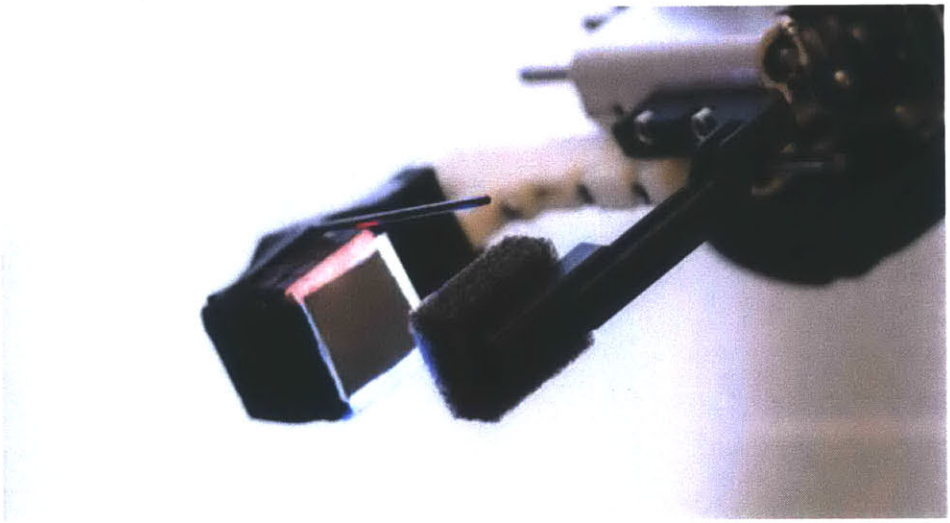


Figure 2-12: Fingertip GelSight sensor mounted on a robot gripper.

- **High resolution**
  - The spatial resolution of the new sensor is 10-40 microns depending on the resolution of the imaging system. Hence it can well sense the human hair that has a diameter of around 100 microns, as well as the printed ink on a ten-dollar note with height less than 100 microns (Figure 2-13).

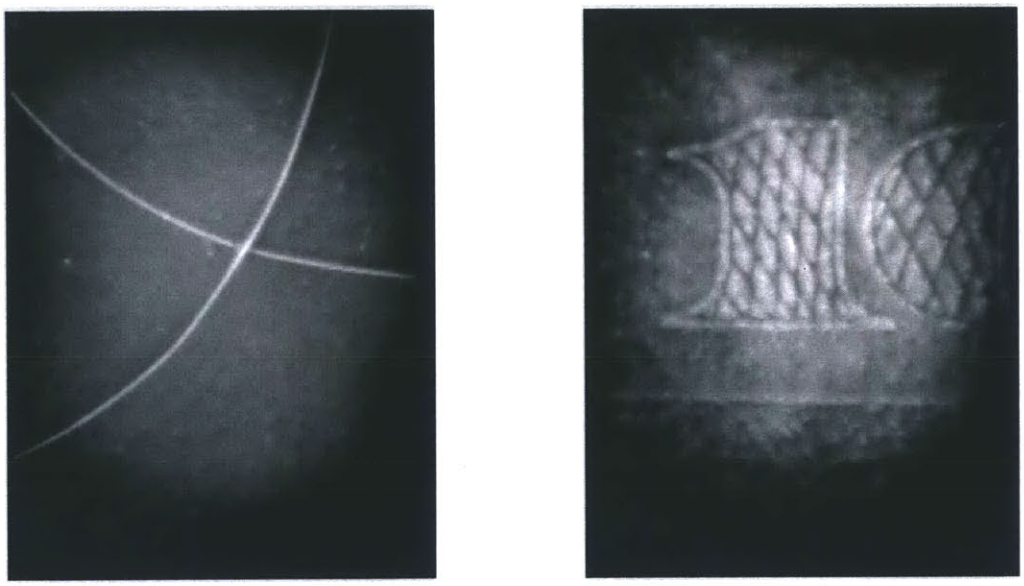


Figure 2-13: 3D images obtained from the fingertip GelSight sensor of two human hairs (left) about 100 microns in diameter and the printed letters on a ten-dollar US note (right) thinner than 100 microns.

- **Compliant**
  - The new sensor itself is soft and compliant, able to sense human fingerprints, as shown in Figure 2-14. Depending on the type of tasks, the sensor can be made of different hardness with different compliance.

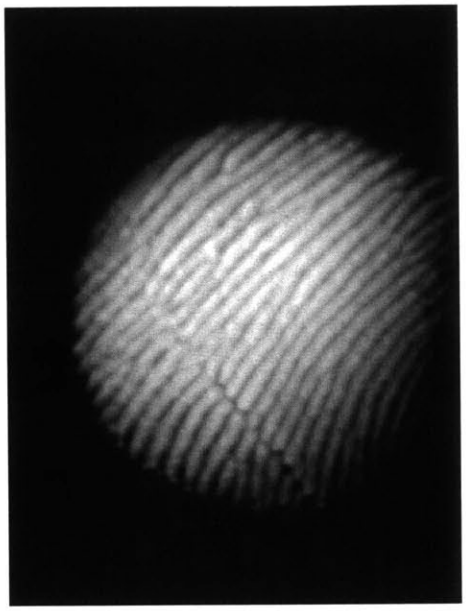


Figure 2-14: 3D image of a fingerprint obtained by the fingertip GelSight sensor.

- **Real-time speed**
  - Currently the sensor can run at a speed of 10-30 frames per second or higher, depending the amount of processing power and image resolution. When we run the algorithm on a 64-bit Windows 7 virtual machine with Intel® Core™ i7-3840QM CPU @ 2.80GHz and 8.8GB RAM, the MATLAB program gives a speed of about 20 frames per second with an image resolution of 320x240. Increasing the processing power of the computer further and/or rewriting the code in C/C++ will further increase the speed at a significant amount which will then be only limited by the speed of the camera, which is 30 frames per second for Logitech C310. Note also that the camera used has a lag of around 100 ms, which may or may not be ignored depending on the applications.

- **Low cost**
    - The sensor is comprised of off-the-shelf components and 3D printed materials: a webcam, 3D printed housing, an elastomer with a thin membrane, acrylics, LEDs and some tapes. Among these the most expensive component is the webcam, which is around $20. The components are listed below in Figure 2-15.

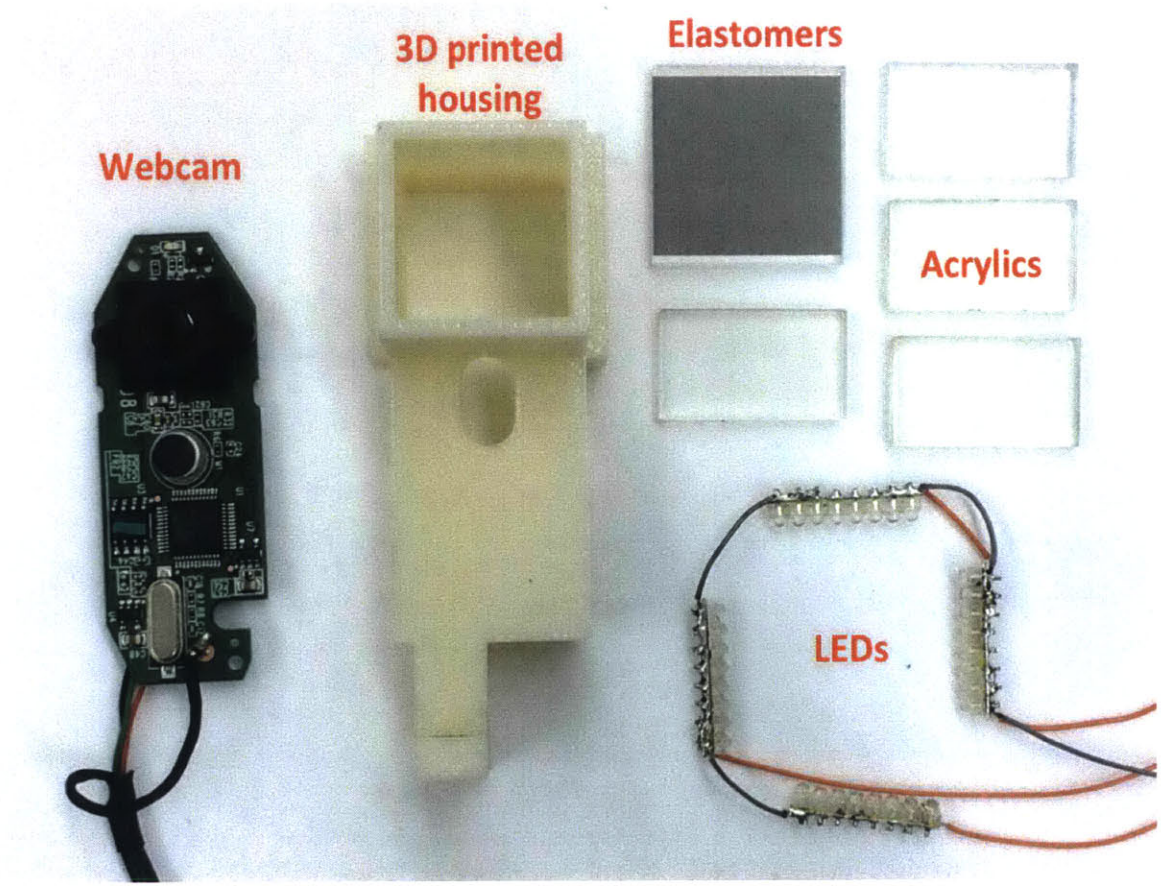


Figure 2-15: Basic components of a fingertip GelSight sensor: webcam, 3D printed housing, elastomer, acrylics, LEDs and circuit.

In summary, the fingertip GelSight sensor is a compact tactile sensor that has high resolution, on the order of tens of microns, and real-time speed, and is compliant and

inexpensive. The special properties of the fingertip GelSight sensor make it suitable for a wide range of tasks which will be discussed in detail in the next three chapters.

Chapter 3

Robotic Fine Manipulation with GelSight

Robotic manipulation and insertion of small parts can be challenging because of the small tolerances typically involved. In this chapter, we present our work using the fingertip GelSight sensor for accurate localization and mapping of the object in hand, and performing insertion tasks with high success rate. This chapter is based on collaborative work with Robert Platt's lab, which was published as Li et al, IROS 2014 [5].

3.1 Introduction

Small parts manipulation and insertion is an important robotics problem that has applications in manufacturing, space and hazardous environments, and medicine. A good example is the problem of grasping and inserting a USB cable, as shown in Figure 1. This insertion is challenging because the tolerances are very low – less than plus or minus one millimeter. Nevertheless, this type of fine manipulation problem is important. Today,

human factory workers are often employed to perform fine insertions and manipulation of exactly this kind.

There are two main types of approaches to performing an insertion such as the USB insertion in Figure 1. The first is force-based perception and/or control. A good example of this is the remote center of compliance (RCC) [62]. In RCC, an active or passive compliance mechanism is developed that causes the peg to slide into the hole when a downward force is applied. Alternatively, it is possible to perceive hole location based on sensed forces [63]. However, this technique is hard to use in many robot manipulation problems because of secondary load paths. Another important approach to performing fine insertions is visual servoing [64], [65]. Here, the robot vision system localizes features or fiducials both on the part to be inserted and on the mating surface. The transform between the two features is estimated using projective geometry in the reference frame of the camera and a robot motion that reduces the error relative to a desired transform is calculated. The advantage of this approach is accuracy: visual servoing can guide part insertion to within ten microns of error [65], [66]. However, it is necessary for the insertion operation to remain within view of the camera during the entire operation (a challenging requirement because the robot hand tends to occlude the mating surface). Moreover, it is necessary to be able to localize features on both the part and the mating surface consistently through time. Since this can often be challenging to do using "natural" features, fiducials are often affixed to the parts in order to facilitate consistent localization [66] (an undesirable modification of the environment).

Instead, this thesis explores a tactile-sensing based approach to the problem. In contrast to force sensing methods, our approach is to use tactile sensing to localize a part accurately relative to the gripper holding it. We assume that the main challenge is localizing the part in the robot hand and not localizing the mating surface. This assumption is reasonable in many scenarios where the mating surface is fixed and can be localized prior to insertion or manipulation. In this case, it is the pose of the part in the hand that is hard to estimate. The key feature of our approach is the use of a tactile map [67]. The tactile map is a model of what the object surface is expected to feel like as a function of contact configuration and is created prior to manipulation (see Figure 3-2 (b)). During manipulation, the robot perceives tactile information regarding the pose of the

object in the hand. By registering this tactile information back to the tactile map, the robot can localize the pose of object relative to the gripper. In this paper, we use a recently developed tactile sensor, known as GelSight [1]. The GelSight sensor reconstructs the 3D geometry of the surface of a contacting object using photometric stereo algorithms. The resolution of the resulting height map is on the order of the number of camera pixels – 320 x 240 in our case. We use feature-based RANSAC operating on this height map both to create tactile maps and to localize a given set of tactile information within a map.

3.1.1 Related Work

The problem of localizing an object using tactile sensing has been studied for a long time. Early work included approaches based on fitting a parameterizable object model to contact points [68], using observability theory to estimate object pose [69], and using active tactile interaction to explore objects [70]. More recent work uses Bayesian estimation. For example, Chhatpar and Branicky use particle filtering to localize a peg with respect to a hole [71]. Gadeyne and Bruyninckx use Markov localization to localize the 3-dof pose of an object [72]. Petrovskaya et al. localize the 6-dof pose of an arbitrary polyhedral object by making a series of surface probes [73]. Corcoran and Platt localize the 6-dof pose of an object held in a whole-hand grasp based on a set of binary contact information [74]. A couple of prior works incorporate the idea of a tactile map. Platt et al. use tactile maps to localize distinctive haptic features in soft materials [76]. Pezzementi et al. use tactile models in order to classify and localize objects (for example, raised letters from a children's play set) [75]. Another important area of related work has to do with other tactile sensors that measure deformation in a deformable membrane. For example, Hristu, Ferrier, and Brockett proposed a deformable membrane tactile sensor that operates by tracking dots printed on a deformable membrane and reconstructing the contact geometry using a finite elements approach [76]. Wettels, Smith, Santos, and Loeb, developed a sensor that measured pressure in a weakly conductive fluid fingertip at a small set of locations [77]. Torres-Jara et al. developed a tactile sensor that used hall effect sensors to measure membrane deformations [78]. An interesting review of human and robot approach to tactile sensing can be found in [79].

3.2 Experiment Setup

The key to robust control of small parts manipulation interactions is accurate tracking and control of the parts involved. Typically, this is accomplished using visual servoing or force-based control. However, these approaches have drawbacks. Instead, we propose a new approach that uses the fingertip GelSight sensor to accurately localize the pose of a part grasped in the robot hand and perform insertion tasks accordingly. Using a feature-based matching technique in conjunction with our fingertip GelSight sensor that has much higher resolution than competing methods, we synthesize high-resolution height maps of object surfaces. For example the Takktile array sensor that was recently supplied as Government Furnished Equipment to Tack B teams in the DARPA Robotics Challenge can sense independent forces over an 8×5 grid with approximately 6 mm resolution [66]. Similarly, the RoboTouch sensor from Pressure Profile systems that is distributed with the Willow Garage PR2 robot has only 24 independent sensor elements on each sensor pad with 6×6 mm resolution [67]. In contrast, the GelSight technology can sense contact geometry at approximately pixel resolution – 320×240 in our current sensors. This is approximately 1,000 times higher resolution than competing sensors.

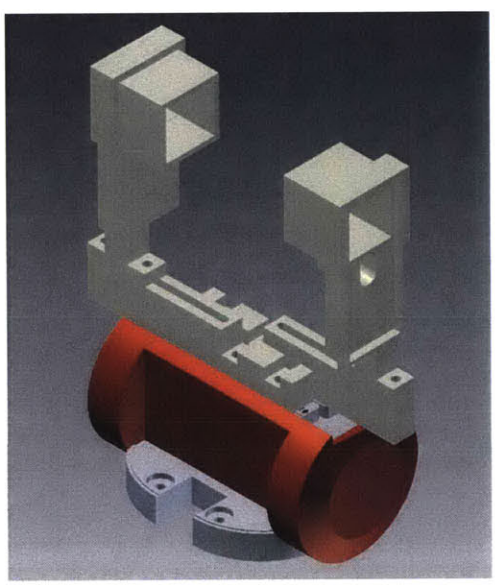
As a result of these high-resolution object height maps, we are able to localize small parts held in a robot hand very accurately. We quantify localization accuracy in benchtop experiments and experimentally demonstrate the practicality of the approach in the context of a small parts insertion problem.

We performed manipulation tasks with a Baxter robot produced by Rethink Robotics, Inc. together with the fingertip GelSight sensor presented above. Figure 3-1 shows how the sensor is integrated and mounted into the Baxter gripper. We have designed two fingers that can each accommodate one sensor. These fingers were manufactured using fused deposition modeling (FDM), a type of 3-D printing. The box at the end of each finger accommodates the illumination apparatus and the camera. The back of each finger has a slot specifically designed for mounting of the camera board from the Logitech webcam (used above). As Fig 3-1(b) shows, only one of the fingers was equipped with a sensor. The other finger opposed the sensor with a compliant mesh pad. Although using only one sensor reduces the amount of information available, we found that having the mesh pad improved the measurements obtained using the one sensor because it pushed

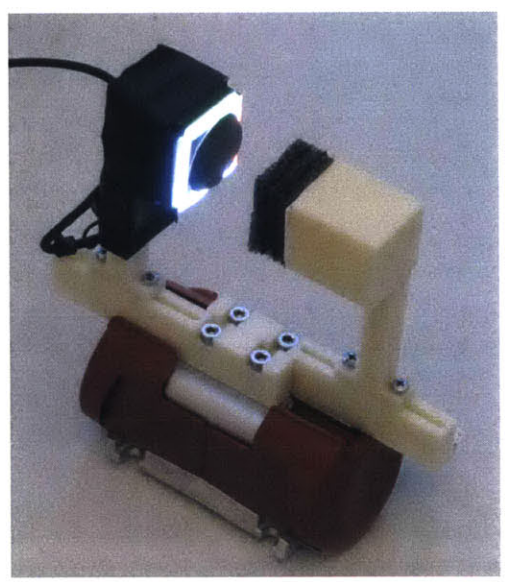
the grasped part into good contact with the one sensor. We have found that the shape of the elastomer gel on the sensor is important. We have explored two alternatives: gel with a flat membrane and gel with a domed membrane. While the flat membrane can make contact over a larger surface area, it can fail when the sensor is not aligned parallel to the object surface. Instead, we have used the domed membrane in our experiments (Figure 2-9).

3.3 Localization and Mapping via Image Registration

The key challenge in using tactile sensing to localize an object held in the robot hand is the creation and use of the *object map*. The object map is a model of what the robot expects to feel as a function of the position and orientation of the object relative to the sensor. The map enables the robot to localize a grasped object in its grip. For example, Figure 3-1(b) illustrates an object map of one side of a USB connector. When the robot grasps the connector (Figure 3-1(a)), the GelSight sensor mounted on its fingertip measures a height map of the portion of the connector surface where it is gripped. By matching this height map with the corresponding portion of the object map, it is possible to localize the grasped object with respect to the gripper.

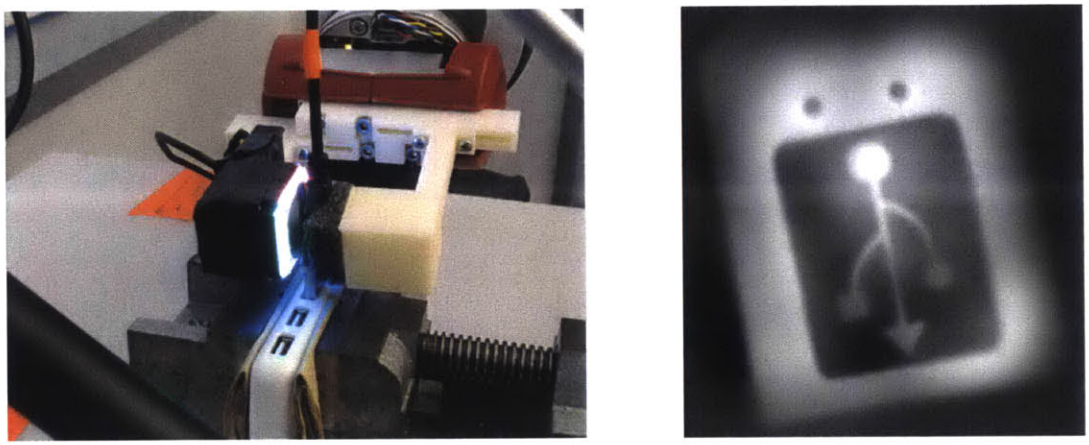


(a)



(b)

Figure 3-1: Integration of the GelSight sensor into the Rethink Robotics Baxter hand. (a) CAD model of the Baxter hand with two fingers. (b) One finger was equipped with a GelSight sensor in our experiments.



(a) (b)

Figure 3-2: The robot, Baxter, inserts a USB connector into a mating hole. The robot relied on tactile sensing in order to localize the connector in the gripper and therefore to calculate how to move the hand. (b) Illustration of the object map that describes what the connector "feels like" to the robot. The object map is used to localize the connector relative to the gripper.

3.3.1 Registration of A Tactile Image

In order to create a new object map or to localize a tactile measurement within a map, it is necessary to register one height map with respect to another. This is very similar to the well-known image mosaicing problem. However, in our case, we are mosaicing height maps rather than RGB images. Nevertheless, we have found that standard feature-based matching techniques can work well. In our scenario, it can be assumed that the two height maps will have nearly the same scale and that there will be no out-of-plane rotation. Therefore, the problem reduces to that of estimating the isometry between the two height maps. Our approach is as follows. First, we localize keypoints and descriptors in both images using a recently developed detection algorithm that locates robust keypoints with binary descriptors, known as BRISK [80]. BRISK is conceptually similar to other feature detectors such as SIFT and SURF, but it is much faster: feature detection runs

approximately one order of magnitude faster than SURF. Also, we hypothesize that binary features are better suited for height maps compared with standard real-valued descriptors. Matches between corresponding features are calculated by thresholding the Hamming distance between descriptors.

We calculate the best-fit pose using RANSAC [81]. Hypothesis poses are sampled uniformly at random by sampling two pairs of matching keypoints. The two keypoint pairs give us a candidate translation, $t \in R^2$, and rotation, $R \in SO(2)$. These are combined to give us a candidate isometry,

$$H = \begin{bmatrix} R & \boldsymbol{t} \\ \mathbf{0} & 1 \end{bmatrix} \tag{1}$$

where $\mathbf{0} = (0, 0)$. For each candidate isometry, we evaluate the Euclidean distance between the transformed points from one height map and the corresponding points from the other. Pairs with a below-threshold distance are labeled inliers. After performing several rounds of sampling, we choose the isometry with the largest number of inliers and evaluate quality of fit. We calculate the least-squares homography, H, between inliers in one height map and inliers in the other. Because we are matching tactile sensor information, the best fit homography should be an isometry (it should have only translation and rotation components). We evaluate the "distance" of the homography to an isometry by evaluating the determinant of the rotation component of the homography. We treat the determinant as a measure of our confidence that the match is correct. Let R be the rotation component of the homography. Then our confidence measure is:

$$c = \max(1 - |1 - \det(R)|, 0). \tag{2}$$

Confidence is highest when $c = 1$. A typical result of this approach to tactile image registration is shown in Figure 3-3. Figure 3-3(a) and (b) show two tactile images of overlapping areas of a penny. Figure 3-3(c) shows the composite registered tactile image.

3.3.2 Mapping

An object map is a height map of a surface of an object. In this work, we focus on

mapping only a single face or side of an object. Some objects, such as a key or a USB connector, are well modeled this way because they are nearly always grasped with one finger on each of the two large flat sides. The map is created on-line in a sequential way. We start by capturing a single height map of some part of the object surface as the "root" of the map. Then, we obtain additional tactile images by touching the object surface in different configurations. Each time a new image is acquired, we attempt to match it to the object map. If the match confidence exceeds a threshold (0.98 in our experiments), then we add it to the map. Height values in areas where the new image overlaps with the current map are averaged with height values from other images. In new areas, the height value is placed on the map directly. Figure 3-4 shows a complete height map of a penny.

3.3.3 Localization Experiments

We performed experiments to characterize the localization accuracy of our approach. The goal of localization is to locate the grasped object relative to the gripper. When the gripper grasps an object, the GelSight sensor captures a height map of a segment of the object surface. This height map is registered with the object map and used to localize the object. Figure 3-5 (a) shows the experimental setup. An object was fixtured to an adjustable x–y–θ platform and the tactile sensor was fixtured in a jig above the platform.

(a) (b) (c)

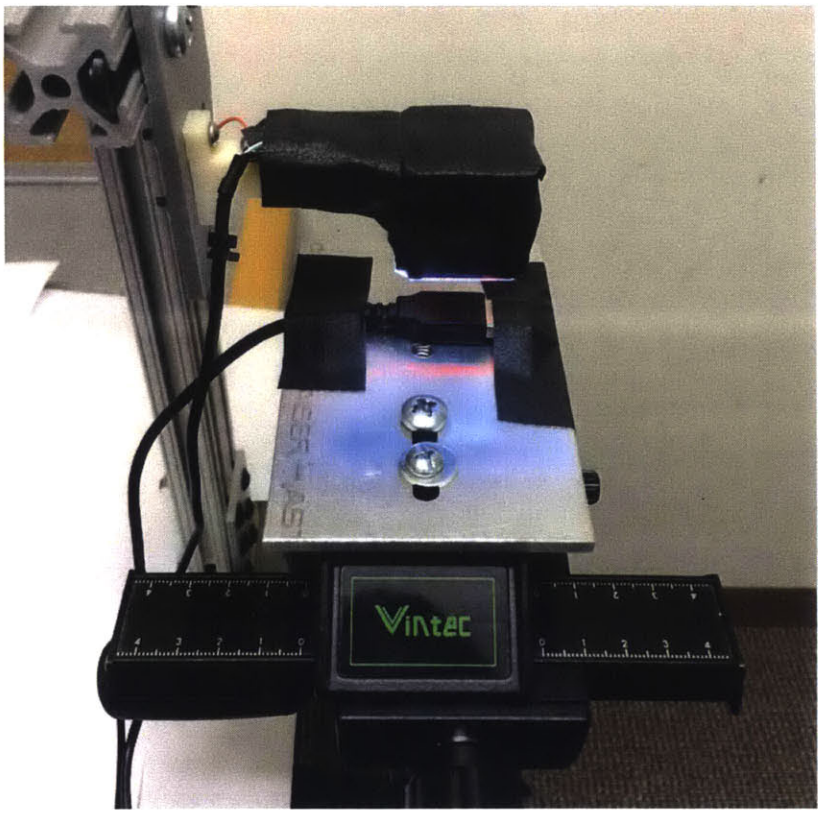
Figure 3-3: (a) and (b) Two height maps created by touching different parts of a penny with GelSight tactile sensor. (c) A composite height map created by registering the two components.

Figure 3-4: Object height map of a penny created by registering several partial tactile images into a single height map.

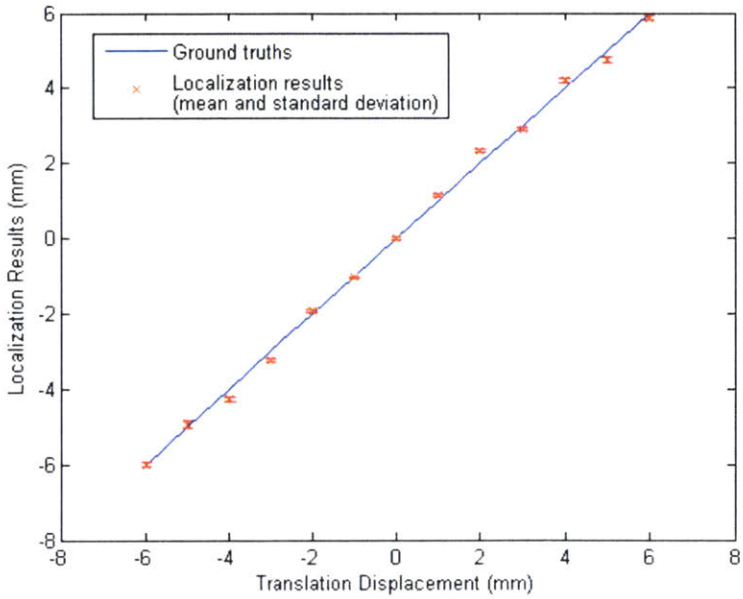
This allowed us to adjust the position of the object relative to the sensor in a controlled way while capturing tactile data. During the experiment, we moved the object to a series of measured positions and orientations relative to the sensor and captured a height map from the sensor. For each height map captured this way, we registered it with respect to the object map (created prior to the experiment) and thereby calculated the pose of the object relative to the sensor. By comparing the measured pose of the object relative to the sensor and the estimated pose based on the registered height map, we were able to calculate localization error statistics.

We evaluated the accuracy of orientation and translation estimates separately. In order to evaluate orientation error, we collected data using a USB connector as the object. The connector was placed at orientations between −90 and +90 degrees in steps of 10 degrees. The comparison between true orientation and estimated orientation is shown in Figure 3-5(b). The mean absolute error was found to be 1.15 degrees with average standard deviation 1.02 degrees. We performed a similar experiment to evaluate translational accuracy (see Figure 3-5(c)). Here, we performed experiments using a quarter placed at displacements between −6 mm and +6 mm with steps of 1 mm. The mean absolute error for translation localization was found to be 0.14 mm, and the average standard deviation 0.055 mm. It is likely that a portion of the translation and orientation error that we report is a result of experimental error related to manual adjustment of the

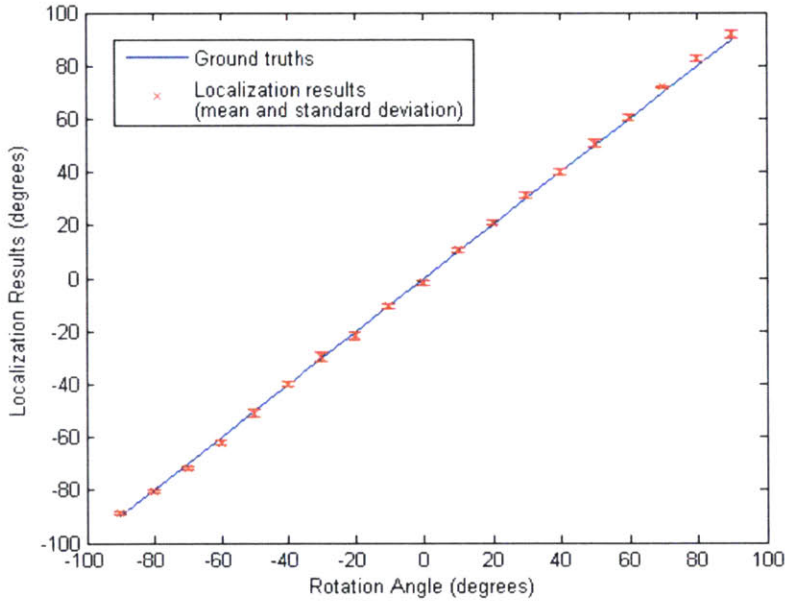
jig to produce the object translation and rotation.



(a)



(b)



(c)

Figure 3-5: Experiments characterizing the localization accuracy of our method. (a) Experimental setup. The GelSight sensor is fixed on a beam facing downwards, and the tripod and bench for hosting the USB with rotation and translation flexibilities. (b) Estimated translation as a function of true translation. (c) Estimated orientation (from tactile data) as a function of true orientation.

3.4 Robot Insertion Experiments

These experiments evaluate the effectiveness of using our approach to small part localization and manipulation in the context of an insertion task.

3.4.1 Setup

The basic setup is as follows. A USB cable hangs from a jig positioned in the robot workspace. The location of the jig is known, but the pose of the USB connector itself varies because the way in which the connector hangs is unknown (see Figure 3-6 (a)). The objective is for the robot to grasp the connector and insert it into a USB mating hole located in a known pose. In order to accomplish this, it is necessary to localize the connector relative to the mating hole with an error of less than approximately ±1 mm. The robot calculates the expected location of the connector based on the jig location. It

reaches to that position and closes the fingers. If the robot cannot localize the grasped connector in its grip, then it releases the hand and moves a small distance and tries again (according to a simple "blind" search procedure). Once the connector has been grasped in such a way that it is localized, then, the robot proceeds with the insertion (see Figures 3-5(b) and (c)).

3.4.2 Connector Alignment

After grasping the connector, the robot moves the gripper to a pose a few centimeters above the mating hole. Then, after localizing the connector in the grasp, it calculates a target pose that will align the connector just above the hole. This occurs as follows. Since we have assumed that the mating hole is fixtured to the environment (*i.e.* the base frame), we can calculate the target transform, ${}^{b}T_{o}^{*}$ for the connector with respect to the base frame. This target transform denotes the pose of the connector hovering just over the hole. The pose of the gripper in the base frame, ${}^{b}T_{g}$, is available using the forward kinematics of the robot. The transform, ${}^{g}T_{m}$, denotes the pose of the map with respect to the gripper. This must be measured during map creation. The map is created by registering tactile images relative to a single root image. The pose of this root image with respect to the gripper must be measured and stored as ${}^{g}T_{m}$. Finally, ${}^{m}T_{o}$ denotes the pose of the object in the map frame. This transform is calculated using feature-based height map registration. Given all of these transforms, we can calculate the target gripper pose as follows. The transforms are related by:

$$ {}^{b}T_{o} = {}^{b}T_{g}\,{}^{g}T_{m}\,{}^{m}T_{o} \,. \tag{3} $$

Given the target transform, ${}^{b}T_{o}{}^{*}$, we calculate

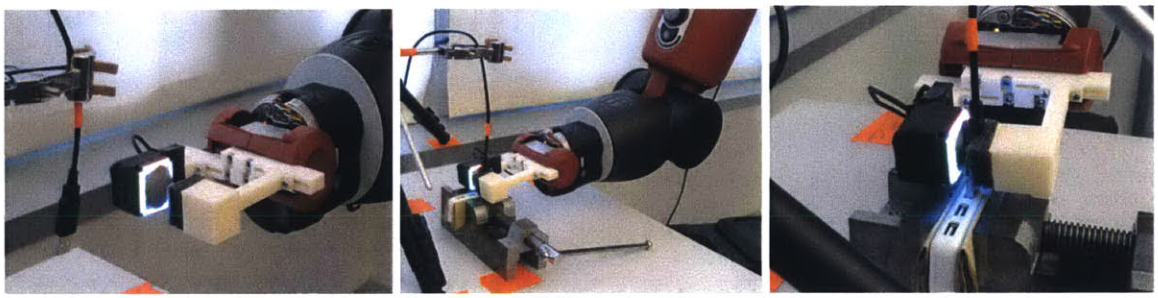
$$ {}^{b}T_{g}{}^{*} = {}^{b}T_{o}{}^{*}\,({}^{m}T)^{-1}({}^{g}T)^{-1} . \tag{4} $$

${}^{b}T_{g}{}^{*}$ describes the desired configuration of the gripper in the robot base frame and is used as input to an inverse kinematics solver or a Cartesian controller.

3.4.3 Connector Insertion

After localizing the USB connector in the hand, we solve Eqn. 4 for the target gripper pose, solve the inverse kinematics problem, and move the gripper to the target location. Rather than using the joint position controller that ships with the Baxter SDK, we developed our own joint position controller. We found that the position error integrator implemented by the SDK did not perform well when the hand contacted the environment. Instead, our controller calculates velocity commands using a position control law: $v^{*} = s^{*}(q^{*} - q)/||q^{*} - q||$, where s^{*} denotes the desired joint speed. This control law is accurate without using integration, and we found it to be stable.

After moving the gripper to the target pose, the USB connector was directly above mating hole. At this point, the robot pressed the connector directly down into the hole. Because we wanted to limit the amount of force that we applied through our sensor, we did not require the USB cable to be fully inserted in order to count the insertion as a success. We only required the connector to be completely inside the mating hole so that continued downward pressure would cause the connector to become completely mated (see Figure 3-6(c)).



(a) (b) (c)

Figure 3-6: (a) Grasping the USB connector. (b) Insertion of the connector into the mating hole. (c) Insertion closeup.

3.4.4 Insertion Experiment and Results

Our system performed 36 USB insertions with two failures using the Rethink Robotics Baxter robot. On each trial, the USB cable was placed in the jig as shown in Figure 11(a). The robot reached forward from the same starting pose toward a fixed pose with respect to the jig and closed the gripper. If the system failed to localize the USB cable in its grasp, it opened the gripper and moved 0.75 cm forward and tried again. If that failed, then the system opened the gripper and moved 1.5 cm back. This process repeated until the USB cable was localized. This procedure resulted in the set of 36 relative gripper-connector configurations shown in Figure 3-7. Of these 36 grasps, 34 of the subsequent insertions succeeded. Poses for the successful insertions are shown in black. The two failures are shown in red. We believe that the two failures were caused by an error in joint position control where the Baxter arm did not reach its target pose.

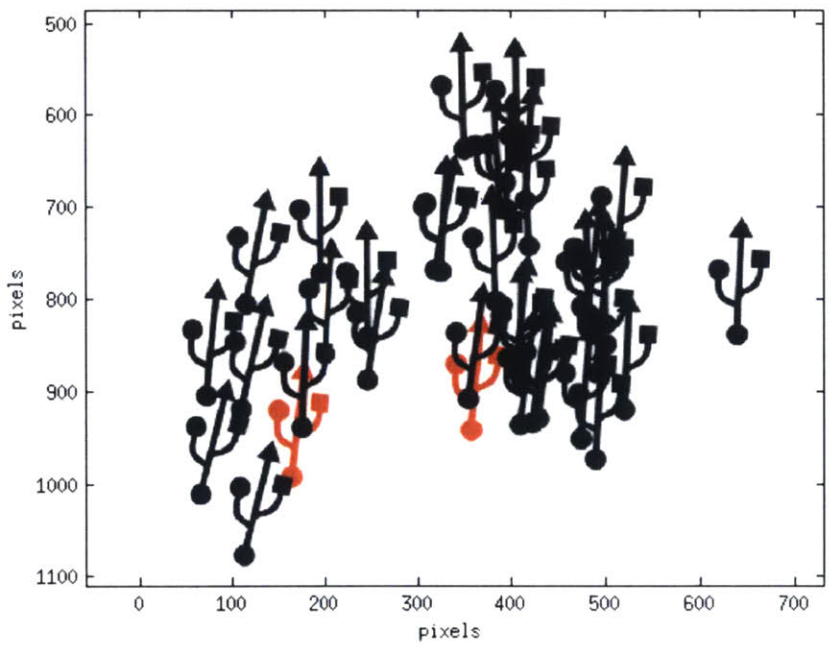


Figure 3-7: The set of 36 gripper-connector poses experienced during our experiments. The poses shown in black were successful. The two in red failed. Each pixel represents 0.005mm of connector displacement.

Fine parts manipulation and/or insertion is very challenging because of the fine tolerances that are typically involved. The key challenge in this kind of task is locating the part in the grasp. The precise pose of the part may be uncertain at the time of grasping

or it may shift in the grasp as the robot moves. In either case, it is typically necessary to re-localize the part precisely just prior to insertion. In this paper, we explore an approach to localizing the part in the hand using tactile information. A key part of this work is our development and use of a robot fingertip version of the GelSight tactile sensor. This sensor delivers height maps of the surface of objects in the grasp at a much finer resolution than what is otherwise available. As a result of this key capability, we are able to use mapping and localization techniques to localize parts in the grasp very accurately.

Chapter 4

Force Estimation and Slip Detection with GelSight

Artificial tactile sensing is still underdeveloped, especially in sensing shear and slip on a contact surface. For a robot hand to manually explore the environment or perform a manipulation task such as grasping, sensing of shear forces and detecting incipient slip is important. In this chapter, we introduce a method of sensing the normal and shear force and detecting slip with a GelSight tactile sensor. Some of the work in this chapter appears in Yuan et al, ICRA 2015 [60].

4.1 Introduction

Tactile sensing is an important aspect of sensation and perception for both humans and robots, because it conveys a great deal of information about the interaction of the body with the environment. Comprehensive reviews of artificial tactile sensing are given in [82] and [83]. Although several successful commercial tactile sensors have been developed (e.g., [84] and [85]), significant limitations exist in currently available robotic

tactile sensors. There is a great need for their continued development, particularly for sensing shear forces and slip at the contact interface. Friction and slip are crucial to dexterous manipulation with soft fingers [86], [87]. Consider the case of a cylinder, which cannot be stably grasped without friction. When a human picks up a cylindrical can of soda, the downward force of gravity is balanced by the upward tangential forces exerted by skin friction at the regions of contact. Likewise, skin friction makes it possible to hold a cylindrical pen during writing, and to exert torque while turning a knob. When friction is insufficient, the result is slip.

4.2 Normal Force Estimation

In a robotic grasping task, the normal force is the squeezing force in the normal direction to the contact surface when holding an object. To a certain extent, deformation of a GelSight elastomer is an indication of the amount of pressure/force applied to it: the larger the pressure, the larger the deformation and hence the larger the height map. In other words, 3D reconstruction of the RGB image may reflect how much pressure or normal force the sensor is experiencing. In fact, when three different spheres of different weights and diameters are put onto the sensor with only gravitational normal force, Figure 4-1 shows the qualitative relationships between the height map and the spheres' weights (or normal forces). Quantitatively speaking, we can see that as the weight becomes larger, the height map also becomes larger.

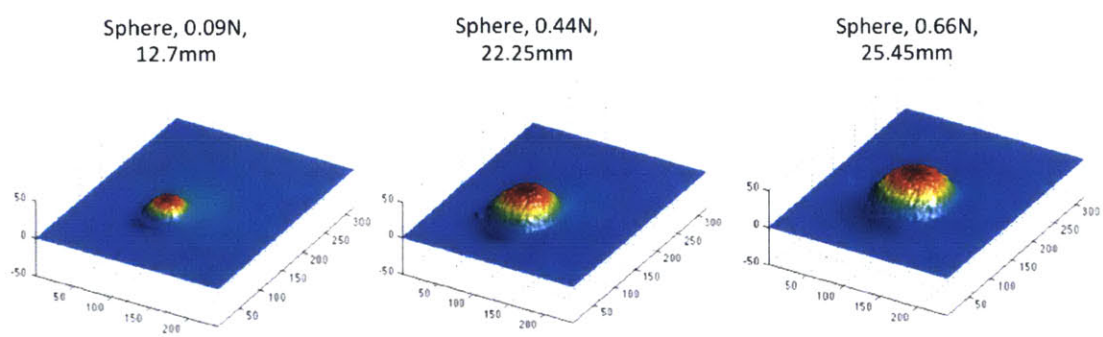
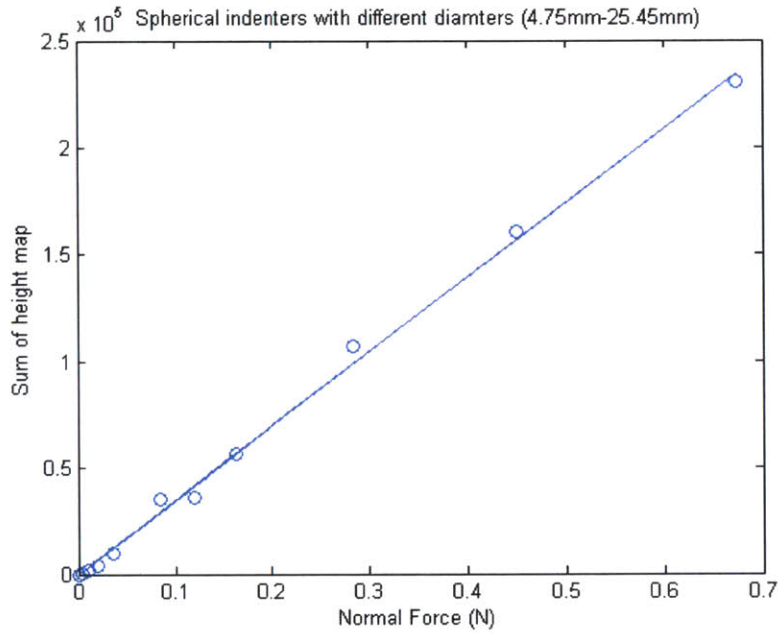
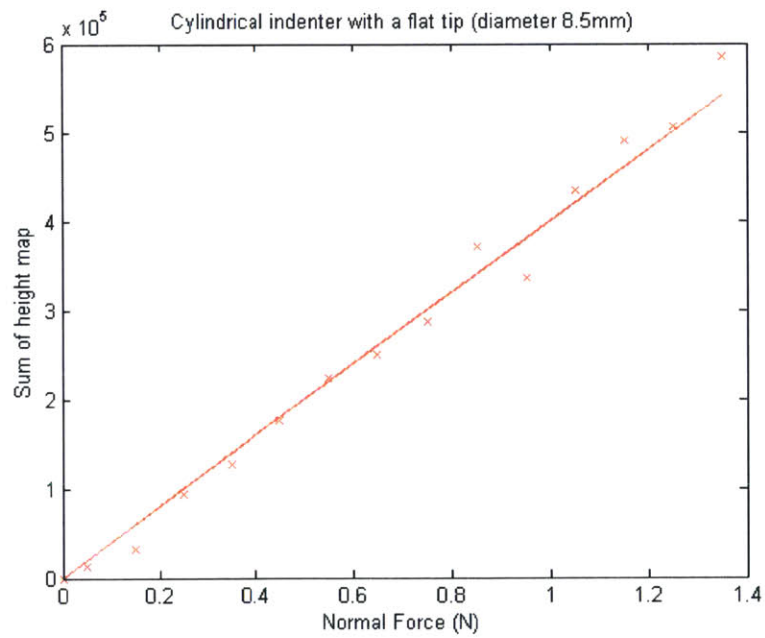


Figure 4-1: Height map vs. normal force for spheres of different weights (0.09N, 0.44N and 0.66N) and diameters (12.7mm, 22.25mm and 25.45mm). Quantitatively speaking, when the weight becomes larger, the height map also becomes larger.



(a)



(b)

Figure 4-2: Plots of sum of height maps vs. normal forces for (a) spheres of different weights and diameters, and (b) a cylindrical indenter with a flat tip that has a diameter of 8.5mm with different weights put on top of it. Here the gravitational force of the indenter is the normal force.

When we experiment with more spheres of different weights and diameters, we obtained the plot as in Figure 4-2(a). It can be seen that there is a strong linear correlation between the sum of height map and the normal force applied onto the sensor. Furthermore, for a cylindrical indenter with a flat tip that has a diameter of 8.5mm, the relationship is also strongly linear, as plotted in Figure 4-2(b). For all experiments, all the contacting areas are within the field of views of the camera. Here you can see that within a certain range (e.g., < 0.7N), given a certain shape, the sum of the height map is linearly proportional the weight itself. The linear factors will depend on the shape of the object and the material properties of the gel. While it is hard to generalize this to a general shaped object and any gel, it still provides meaningful results – for the same shape, the larger the force, the bigger the sum of the height map. Note that for different shapes, the linearity may not be the same largely due to the non-ideal experimental conditions and imperfections of the gels. Also note that for different gels of different softness and thickness, the curve may also look different. Finding the exact relationship between the height map and normal force across different shapes of indenters and elastomers is nontrivial. More experiments and thorough analysis needs to be done in future work for various conditions including different shapes of objects and gels.

4.3 Shear Force Estimation

Besides normal force, shear force is another important parameter for robotic manipulations, such as for stable grasping and insertion. Humans even use shear force to roughly estimate the weight of an object. To estimate the shear force, we placed black markers onto the gel membranes (Figure 4-3) so that we can track displacements of the surface without interfering much with the height map or normal force. The general guideline for placing the markers is that they should be as small as possible while remaining visible under the camera, with a density not affecting much GelSight's original function of getting the surface topography or height map. The average length of the markers used here is around 0.4mm and average distance between two adjacent markers is about 1.2mm. Figure 4-4 shows a version of the fingertip GelSight sensor for shear force estimation (a), the camera view of the GelSight with markers and a finger pressed

onto it applying slight twisting (b), 3D reconstruction or height map obtained that is an indication of the normal force (c), and the corresponding marker displacements (yellow arrows) (d).

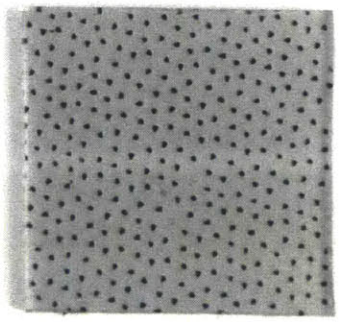


Figure 4-3: Elastomer gel with black markers.

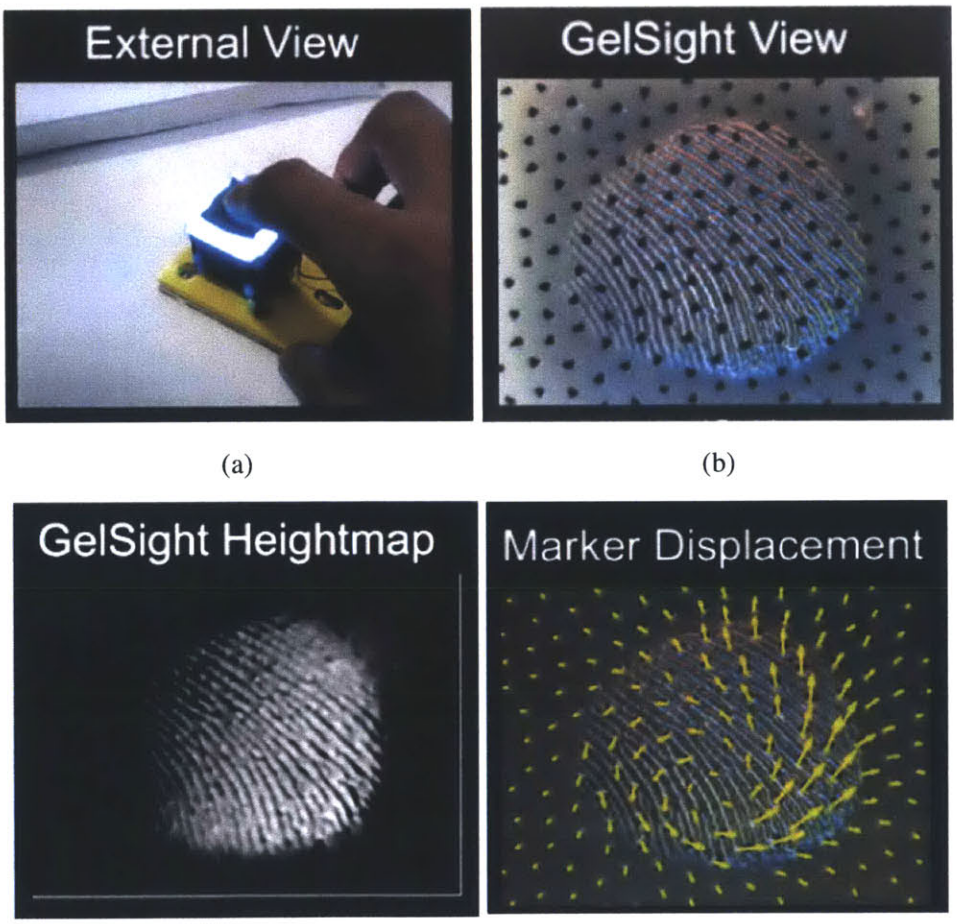


Figure 4-4 [60]: (a) A version of the fingertip GelSight sensor for shear force estimation. (b) The camera view of the GelSight with markers and a finger pressed onto it and applying slight twisting. (c) 3D reconstruction or height map obtained that is an indication of the normal force. (d) Marker displacements (yellow arrows) that indicate the amount of shear force.

The algorithm for tracking the markers is as follows: First, threshold the image and convert it to a binary image. Second, locate blobs of the markers and the centroid of each blob. Third, find the nearest neighbor to each centroid in the subsequent frames and calculate the displacement. On a 64-bit Windows 7 virtual machine with Intel® Core™ i7-3840QM CPU @ 2.80GHz and 8.8GB RAM, the MATLAB program gives a speed of about 25-30 frames per second with an image resolution of 320x240.

In [60], we did experiments to quantify the relationship between the shear force and marker displacement, and found that to be highly linear, as shown in Figure 4-5 and Figure 4-6 (or Fig. 4 and 5 in [60]).

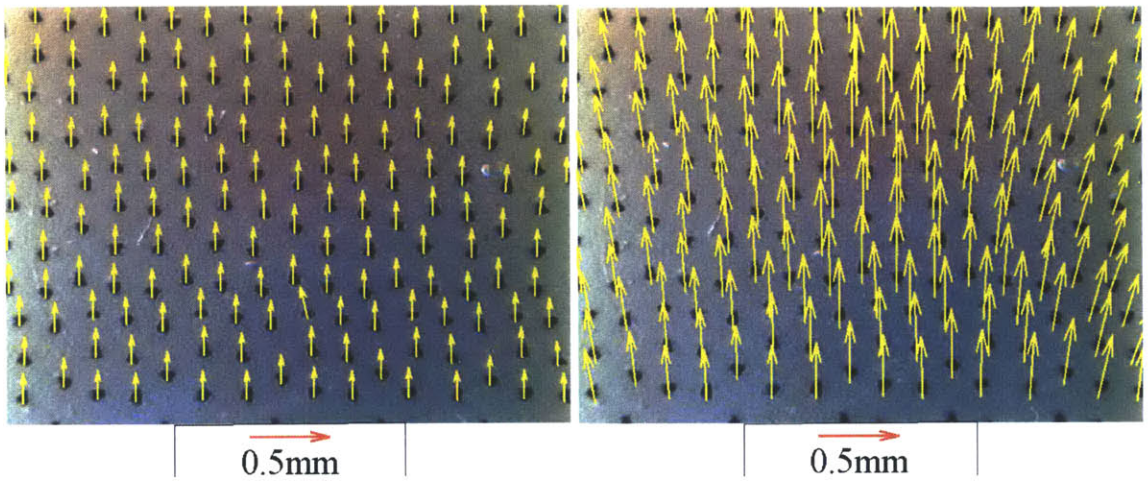


Figure 4-5 [60]: The displacement field for pure shear load. When the load is small (left), the shear displacement field on contact surface is homogenous and in the shear direction; when the load is large, the field is inhomogeneous and directions diverge.

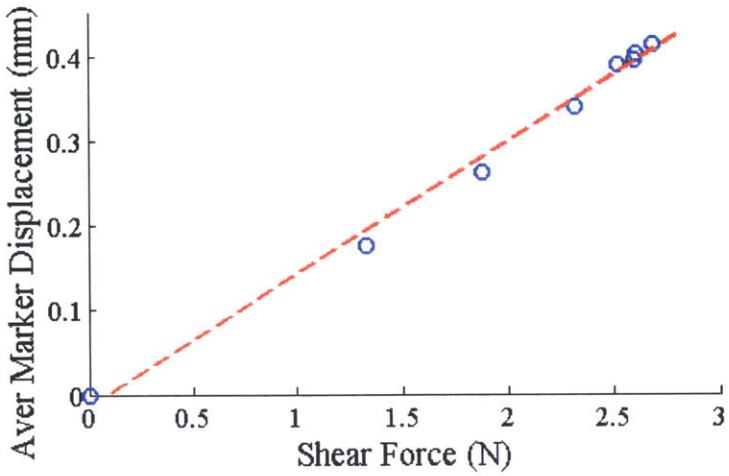


Figure 4-6 [60]: Relationship of the average marker displacements and shear force. All in quasi-static states under the large flat-ended indenter. The shear force is proportional to the average shear displacement of the elastomer over the contact area.

4.4 Slip Detection

We use two parameters for slip detection: force ratio and motion ratio.

4.4.1 Force Ratio

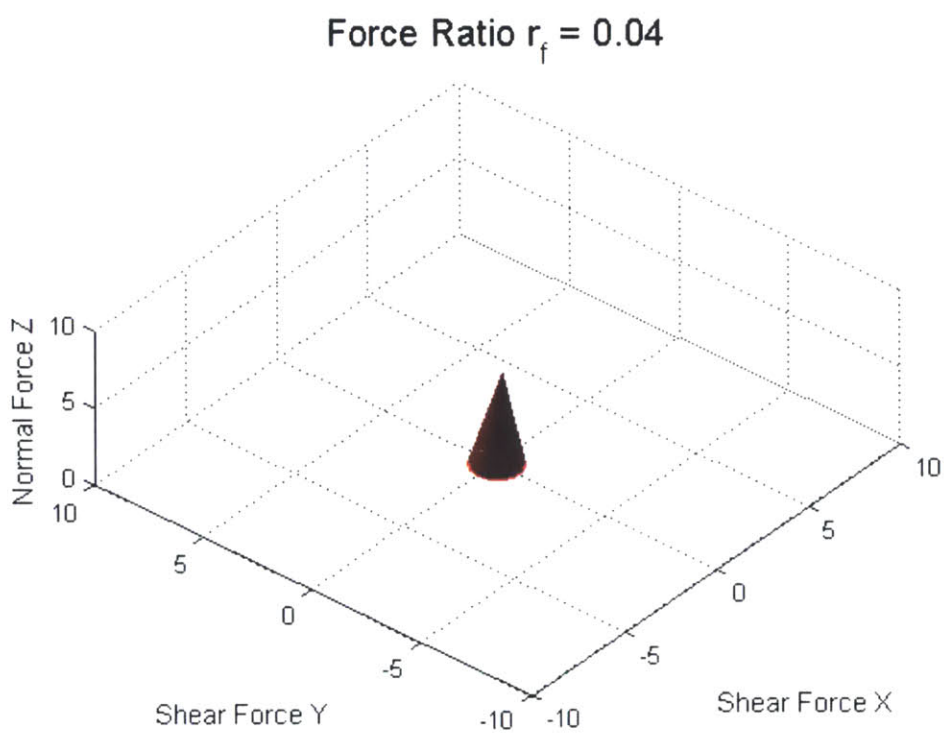
The force ratio, denoted as r_f, is defined as:

$$r_f = \frac{F_s}{F_n}$$

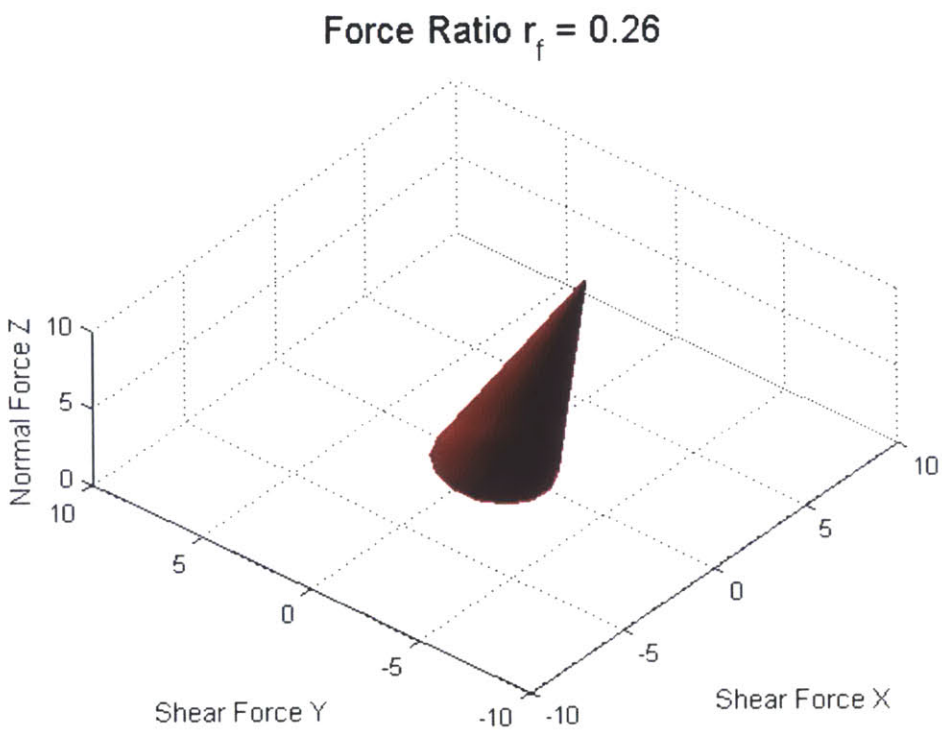
where F_n is the normal force represented by the sum of height map and F_s is the shear force represented by the sum of marker displacement. For surfaces at rest relative to each other, $r_f \leq \mu_s$, where μ_s is the coefficient of static friction which is usually larger than the kinetic counterpart and different for different contact surfaces. For a particular contact surface, slip occurs when $r_f > \mu_s$. For surfaces in relative motion, $r_f = \mu_k$, where μ_k is the coefficient of kinetic friction.

The coefficient of friction (COF) depends on the types of materials used and can only be measured empirically; for example, ice on steel has a low COF while silicone rubber on pavement has a high COF as is the case for GelSight. Through experiments, it is found that when there is relative motion between the surface and the GelSight sensor, the COF

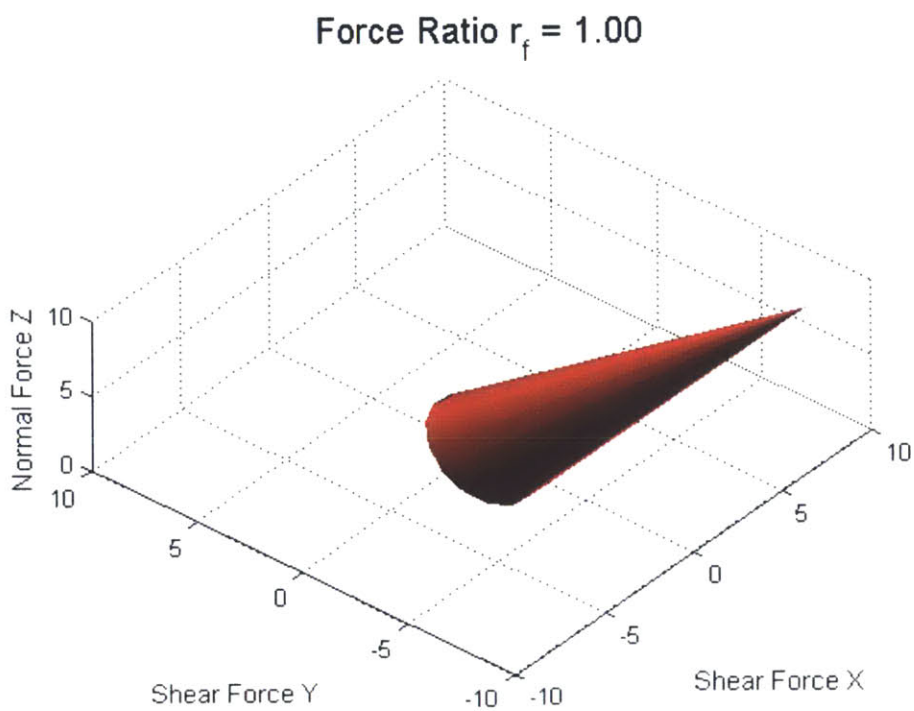
is generally larger than a certain threshold, e.g., 1. This implies that the friction ratio can somehow provide useful information to infer slip. However, that is not the whole story yet because of different COFs for different materials. In some cases, however, depending on the materials used, a COF larger than the threshold does not necessarily imply relative motion or slip. Hence we will need a second parameter to determine there is indeed motion. Figure 4-7 below shows the force ratios under various loads with the normal force and shear force shown in 3D views, where the direction and magnitude of the cone represents the direction and magnitude of the overall force. Note that the force ratio computed depends on the surfaces and the thickness and softness of the gels.



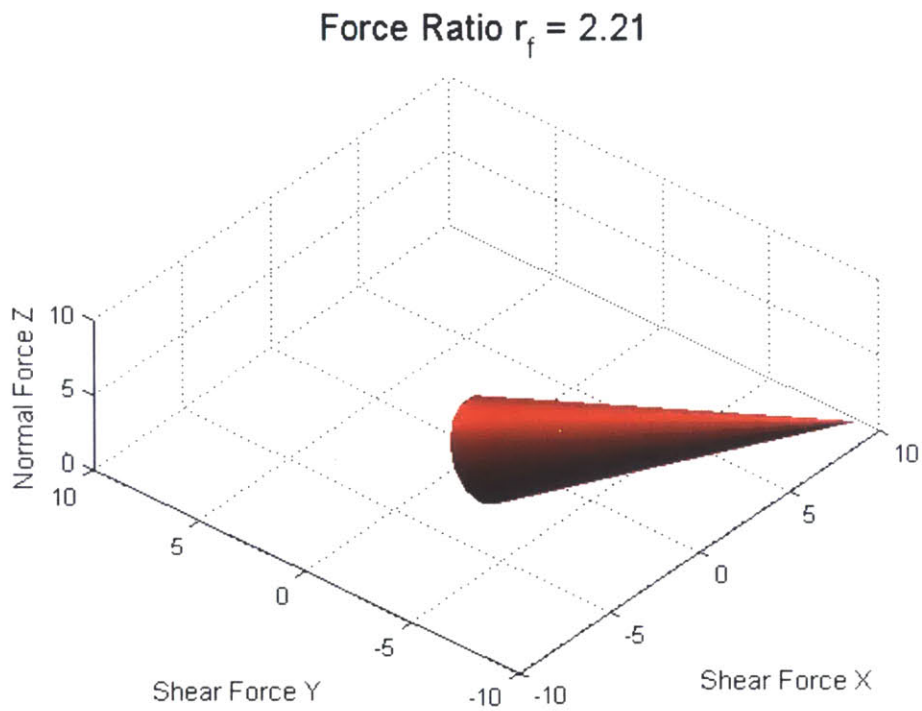
(a)


Force Ratio r_f = 0.26
Normal Force Z
10
5
0
10
5
0
-5
-10
Shear Force Y
-10
-5
0
5
10
Shear Force X

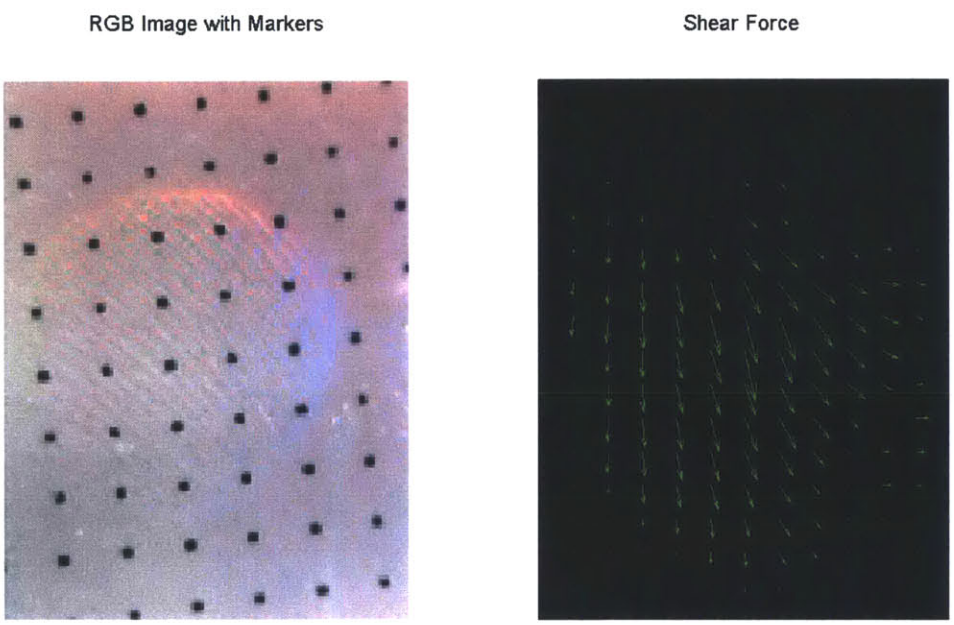
(b)


Force Ratio r_f = 1.00
Normal Force Z
10
5
0
10
5
0
-5
-10
Shear Force Y
-10
-5
0
5
10
Shear Force X

(c)


Force Ratio r_f = 2.21
Normal Force Z
10
5
0
10
5
0
-5
-10
Shear Force Y
-10
-5
0
5
10
Shear Force X

(d)


RGB Image with Markers
Shear Force

(e)

Figure 4-7: (a)-(d) Force ratios under various loads where direction and magnitude of the cone represents the direction and magnitude of the normal and shear force; (e) the RGB image and shear force represented by marker displacements in (d).

4.4.2 Motion Ratio

As discussed above, the second parameter used to infer slip is called motion ratio, denoted as r_m and defined as:

$$r_m = \frac{A_m}{A_c}$$

where A_m is the motion area between two successive image frames, and A_c is the overall contact area.

The contact area can be inferred by first thresholding sum of the RGB channels for the RGB image with the background image subtracted; that is, if the sum is greater than a threshold (e.g., 0.1), we set the pixel value of the corresponding binary image to be 1 and otherwise set it to be 0. We then calculate the number of pixels in the binary image that have a value of 1 and assign that as the contact area. The motion area can be calculated in a similar manner by first subtracting two consecutive image frames to obtain the difference RGB image, and then thresholding sum of the RGB channels for the absolute difference RGB image (e.g., > 0.1) to find the blob area. The reason we use **absolute** difference is that the process should be reversible with the same slip detection results and it should not matter how we subtract the two consecutive images, i.e., image1 – image2 or image2 – image1.

The motion ratio is an indication of the motion status, where $r_m = 0$ means no motion and a larger r_m means partial or overall motion, which may infer the state of incipient, partial or overall slip. Note that due to the way that the motion ratio is calculated, the motion ratio may be larger than 1. One may argue that when the object and the elastomer are moving together, the motion ratio may also be large but it is not true slip. In this case, we call it quasi slip and do not distinguish it from true slip for the purpose of grasping tasks, as in both cases the robot needs to grip tighter to prevent potential or further slip. Figure 4-8 shows how to obtain the contact area and motion area for two successive image frames; here the motion ratio is given by $r_m = \frac{A_m}{A_c} = \frac{1360}{2072} \approx 0.656$.

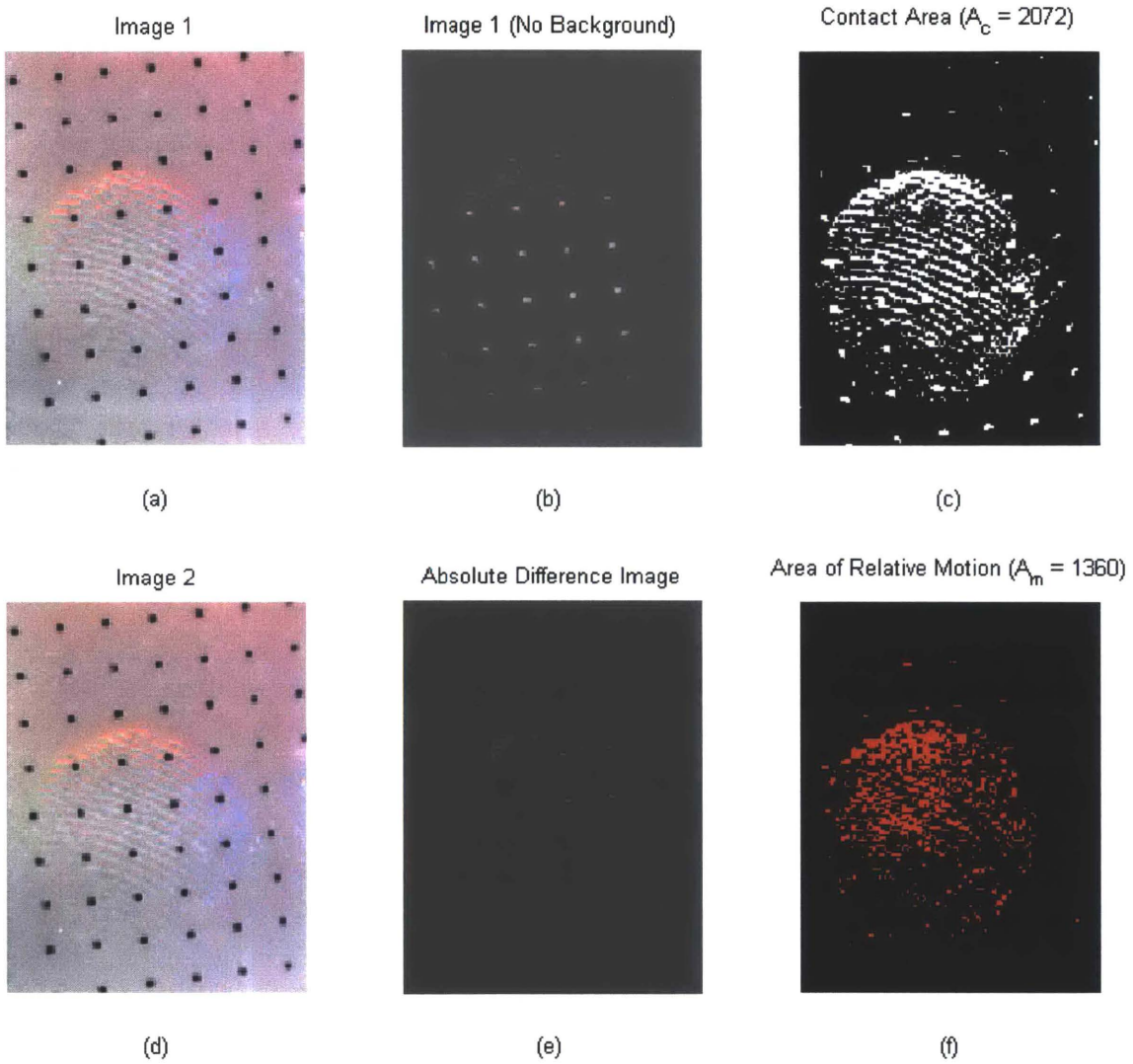


Figure 4-8: (a) RGB Image 1; (b) Image 1 with background subtracted; (c) Binary image obtained by thresholding sum of RGB channels for Image 1, and a threshold of 0.1 is used in this case; (d) RGB Image 2; (e) Absolute difference image between Image 1 and Image 2; (f) Area of relative motion highlighted in red by thresholding the absolute difference image in (e), and a threshold of 0.1 is used in this case.

As discussed in Section 4.4.1, force ratio alone may not fully indicate slip, and motion ratio was introduced to further quantify actual relative motion. On the other hand, motion ratio alone may not fully indicate slip either. One example is in the process when an object is being pressed against the gel, the motion ratio may also be large due to a sudden increase in contact area, but the force ratio is very small because of small shear

force and large normal force. In this case the force ratio eliminates such cases from true slip. In general, we can infer slip by a combination of the force ratio and motion ratio. That is, slip occurs when

$$r_f > T_f \text{ and } r_m > T_m$$

where T_f and T_m are the thresholds for the force ratio and motion ratio, respectively.

Experimentally good thresholds for the fingertip GelSight sensor used for the friction ratio and motion ratio are 1 and 0.5 respectively. That is, we denote slip occurs when $r_f > 1$ and $r_m > 0.5$. Once slip occurs, the robot can take some further actions such as grasping tighter to prevent further slip. Note that smaller thresholds means better sensitivity with a larger possibility of false alarm. For practical applications, a compromise between sensitivity and possibility of false alarm should be made to find proper thresholds.

4.5 Robotic Grasping with GelSight

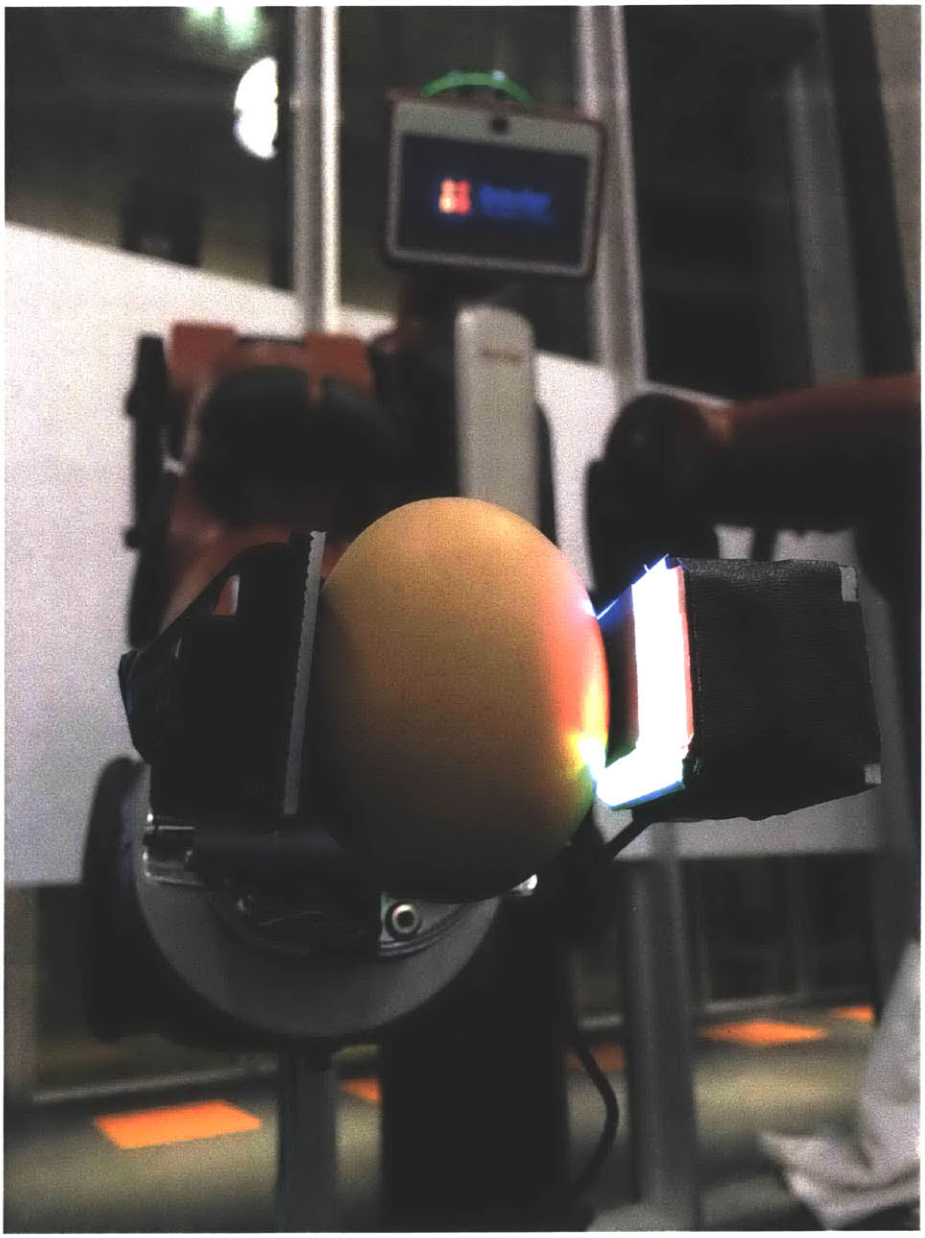
Equipped with the fingertip GelSight sensor, the Baxter robot is now able to sense the normal and shear forces as well as detect slip. For a grasping task, the goal is to grip an object hard enough to resist slip, but at the same time not to crush it. With the force feedback, the robot can take then certain actions accordingly depending on the state of grasping. For example, when it detects incipient slip or full slip, it can close the gripper more to prevent further slip.

The grasping algorithm works in the following way:

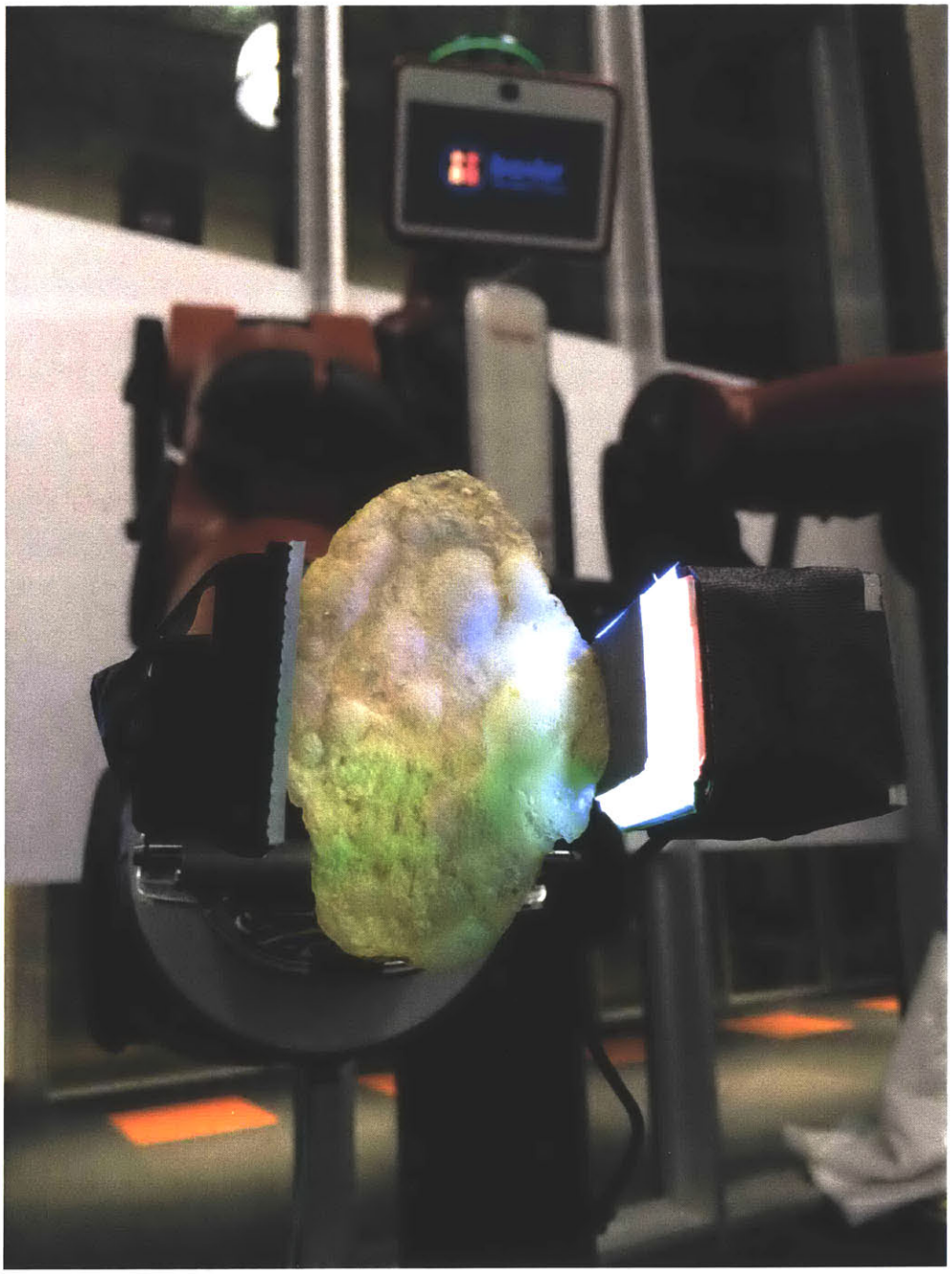
- **Step 1: Initial contact**
    - Keep closing the gripper until it gently touches the object when the maximum of the height map exceeds a threshold, e.g., 2.
- **Step 2: Lifting the object**
    - If slip is detected during the lifting process using the method described in Section 4.4, close the gripper further to grip the object tighter. Otherwise remain the same gripper width.

As a result of the force feedback, the Baxter can now grasp a wide range of objects, ranging from relatively heavy things such as a bottle of water to easily-broken things

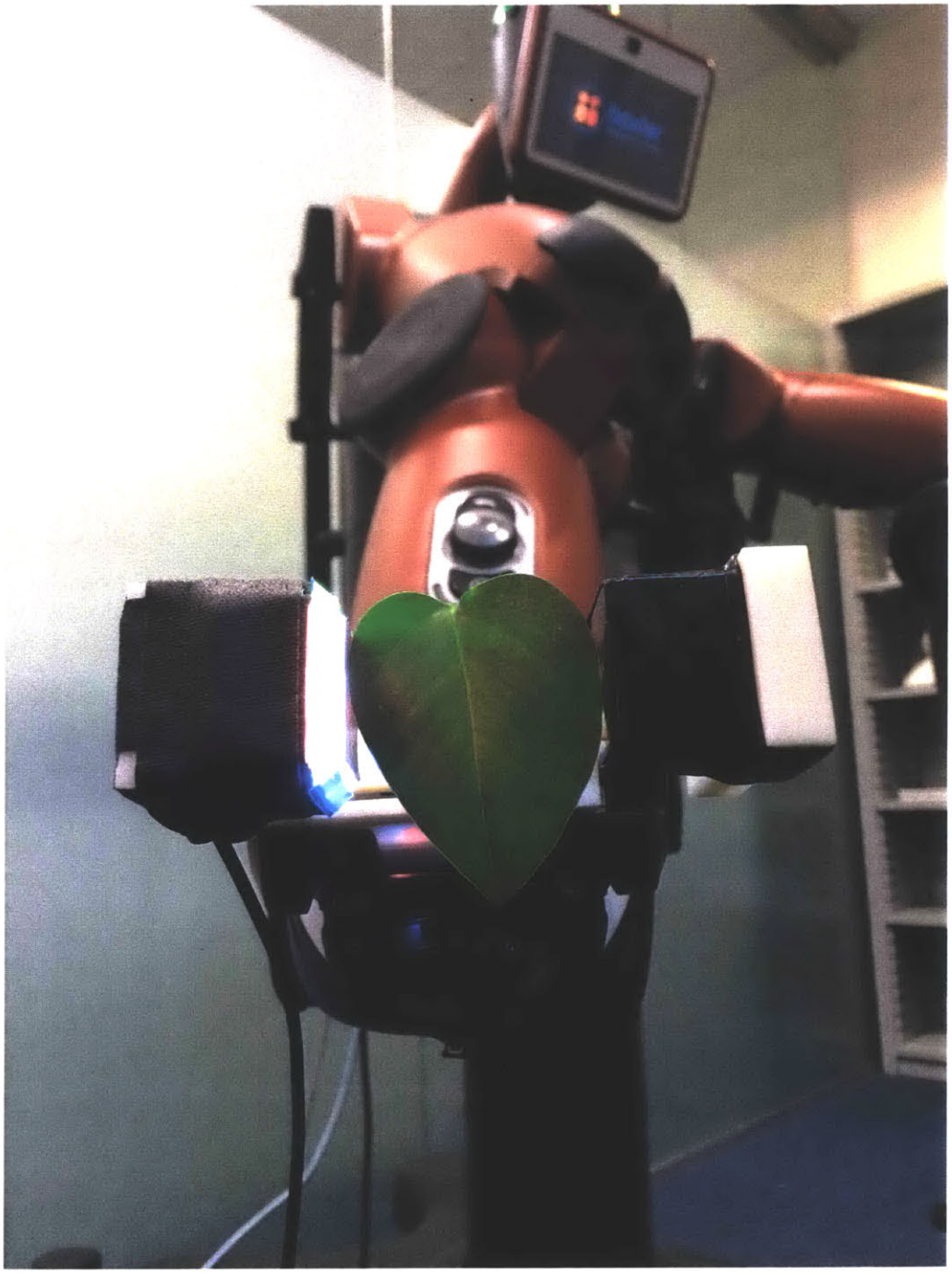
such as eggs, potato chips, and even leaves. Figure 4-9 show some pictures of the grasping in action for different objects without manually changing the parameters in the codes. It is worth mentioning that grasping leaves requires extremely high sensitivity and was not done before with other sensors to the best knowledge of the author. This demonstrates the versatility and huge potential of GelSight for various automated tasks.



(a) Baxter grasping an egg with the use of fingertip GelSight sensor



(b) Baxter grasping a potato chip with the use of fingertip GelSight sensor



(c) Baxter grasping a leaf with the use of the fingertip GelSight sensor

Figure 4-9: Baxter grasping an egg (a), a potato chip (b) and a leaf (c) with the use of the fingertip GelSight sensor that provides real-time force feedback.

Chapter 5

Material Recognition with GelSight

Sensing surface textures by touch is a valuable capability for robots. Until recently it was difficult to build a compliant sensor with high sensitivity and high resolution. The GelSight sensor is compliant and offers sensitivity and resolution exceeding that of the human fingertips. This opens the possibility of measuring and recognizing highly detailed surface textures through touch. The GelSight sensor, when pressed against a surface, delivers a height map. This can be treated as an image, and processed using the tools of visual texture analysis. We have devised a simple yet effective texture recognition system based on local binary patterns, and enhanced it by the use of a multi-scale pyramid and a Hellinger distance metric. We built a database with 40 classes of tactile textures using materials such as fabric, wood, and sandpaper. Our system can correctly categorize materials from this database with high accuracy. This suggests that the GelSight sensor can be useful for material recognition by robots. This chapter details our work in material perception with GelSight tactile sensing, and is based on work in Li and Adelson, CVPR 2013 [3].

5.1 GelSight Image Properties

Understanding surface properties is important for humans and robots, and the sense of touch is an important source of information. Tactile sensing can involve many kinds of information, including temperature, slip, vibration, etc.. In this work we focus on information about local surface geometry, which we will call surface texture here.

A GelSight sensor delivers a detailed height map of the surface being touched, in the form of a function $z(x,y)$, where (x,y) are the point coordinates. This can be treated as an image, and the interpretation can be approached as a vision problem. Figure 5-1 (a) shows a photo of a piece of denim; Figure 5-1(b) shows the height map derived by GelSight, displayed as a surface plot. Figure 5-1(c) shows the height map displayed as a gray image. Figure 5-1(d) – (f) show a photo of a piece of sandpaper, the height map derived by GelSight displayed as a surface plot and the height map displayed as a gray image.

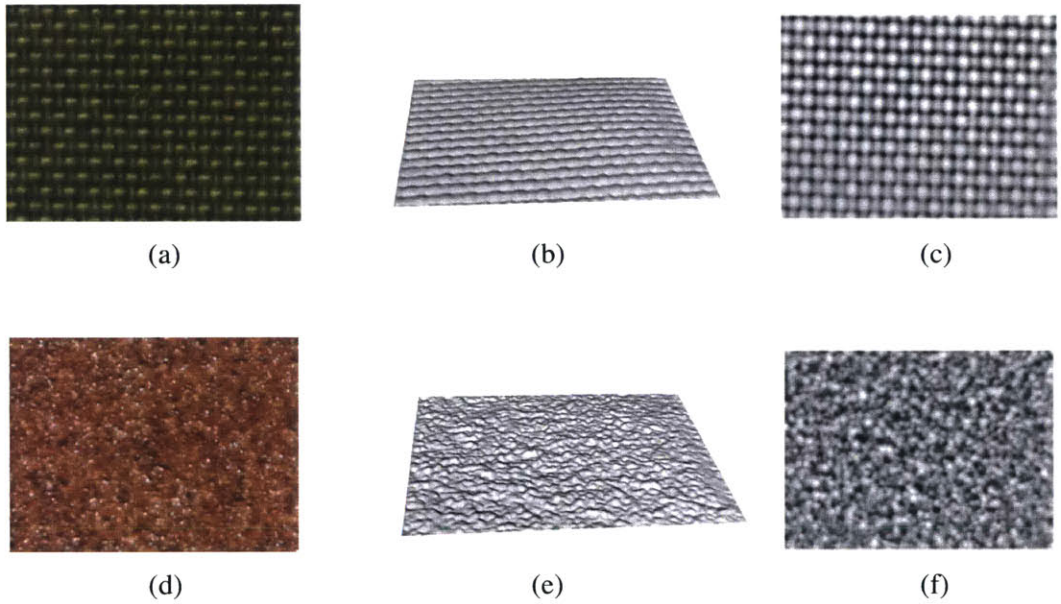


Figure 5-1: (a) A 2-dimensional (2D) photo of a piece of denim. (b) GelSight height map of the denim rendered at a different view. (c) 2D gray image of the denim height map in (b) with brightness corresponding to the height levels. (d) A 2D photo of a piece of sandpaper. (b) GelSight height map of the sandpaper rendered at a different view. (f) 2D gray image of the height map in (e) with brightness corresponding to height levels.

The problem of tactile texture recognition has some specific properties of note. Many vision problems are simplified, since a height map involves no confounds with shading, albedo, distance, etc.. The spatial scale is fixed, since the sensor is in direct contact with the surface and the camera inside the GelSight device is at a fixed orientation and distance with respect to the sensor.

In most cases, the orientation of the texture will be unknown. For example, a denim texture might occur with an arbitrary rotation, so we want a recognition system that is rotationally invariant. The recognition problem is inherently statistical. For example, every patch of 220-grit sandpaper will look different (at the pixel level) than every other patch of 220-grit sandpaper. Thus we are confronted with a classical texture recognition problem. We cannot recognize the sandpaper with a simple template match; rather, there are certain image statistics that will characterize the sandpaper even though each patch is slightly different.

Height maps obtained using GelSight have some special characteristics. To some extent, GelSight images are sensitive to the amount of force applied. Even for the same surface, with slightly different forces, the gray-scale images can have different gray levels due to different levels of deformations of the gel and/or texture surface. Furthermore, the relative orientation between the gel and the texture can be different for each measurement.

5.2 Classification Algorithms

5.2.1 Traditional Local Binary Pattern

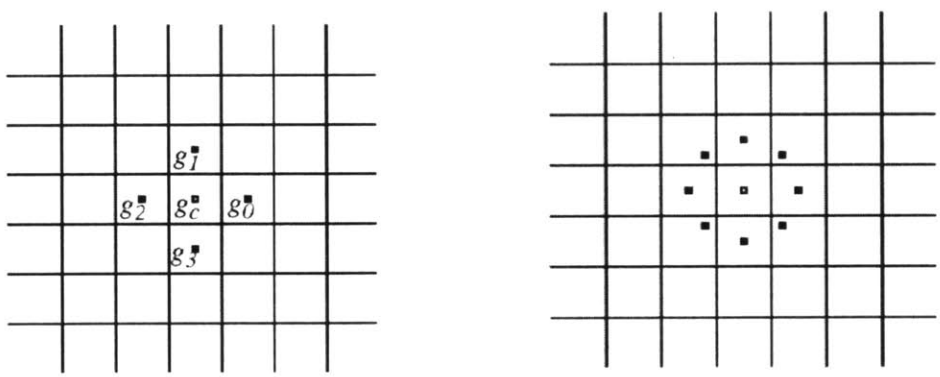
Local binary pattern (LBP) [9] is a texture operator for gray-scale and rotation invariant texture classification. It characterizes local structure of the texture image by considering a small circularly symmetric neighbor set of P members on a circle of radius R. The neighborhood is thresholded at the gray value of the center pixel into a binary pattern, which is then weighted by a binomial factor and summed to obtain the LBP value:

$$LBP_{P,R} = \sum_{p=0}^{P-1} s(g_p - g_c)2^p \tag{5}$$

where

$$s(x) = \begin{cases} 1, & x \geq 0 \\ 0, & x < 0, \end{cases} \tag{6}$$

$LBP_{P,R}$ is the LBP value, g_c is the gray value of the center pixel of the local neighborhood, g_p $(p = 0, \ldots, P-1)$ correspond to the gray values of the P equally spaced pixels on a circle of radius R $(R > 0)$ [9]. If the coordinates of g_c are (0,0), then the coordinates of g_p are $\left(R \cdot cos\left(\frac{2\pi p}{P}\right),\ R \cdot sin\left(\frac{2\pi p}{P}\right)\right)$. The gray values of neighbors that do not fall exactly in the center of pixels are estimated by interpolation. Figure 5-2 shows the neighborhoods with different P and R values.



(P=4,R=1.0) (P=8,R=1.0)

(a) (b)

Figure 5-2 [9]: Illustration of local binary patterns with P equally distributed members on a circular neighborhood of radius R. (a) $P = 4, R = 1$. g_c is the gray value of the center pixel of the local neighborhood, g_p $(p = 0, \ldots, P-1)$ correspond to the gray values of the P equally spaced pixels on a circle of radius R $(R > 0)$. (b) $P = 8, R = 1$.

Signed differences $g_p - g_c$ are not affected by changes in monotonic changes in gray values of pixels; hence LBP is invariant to monotonic gray-scale shifts. Rotation

invariance is achieved by assigning a unique identifier $LBP_{P,R}^{ri}$ to each rotation-invariant local binary pattern, i.e.,

$$LBP_{P,R}^{ri} = \min\{ROR(LBP_{P,R}, i) \mid i = 0,1,\dots,P-1\}, \tag{7}$$

where *ROR(x,i)* performs a circular bit-wise right shift on the *P*-bit number *x i* times. The superscript *ri* denotes rotation invariance.

Furthermore, Ojala et al. [9] defined uniformity measure *U* as the number of spatial transitions (bitwise 0/1 or 1/0 changes) in the pattern, and designated "uniform" patterns as those with *U* not more than 2. The rotation invariant "uniform" LBP operator $LBP_{P,R}^{riu2}$ is defined as:

$$LBP_{P,R}^{riu2} = \begin{cases} \sum_{p=0}^{P-1} s(g_p - g_c) & if\ U(LBP_{P,R}) \le 2 \\ P+1, & otherwise, \end{cases} \tag{8}$$

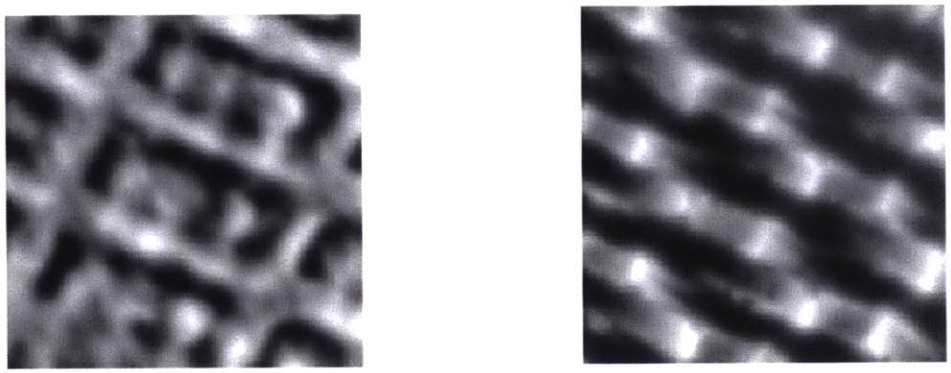
where

$$\begin{aligned} U(LBP_{P,R}) = {} & |s(g_{P-1} - g_c) - s(g_0 - g_c)| \\ & + \sum_{p=1}^{P-1} |s(g_p - g_c) - s(g_{p-1} - g_c)| \end{aligned} \tag{9}$$

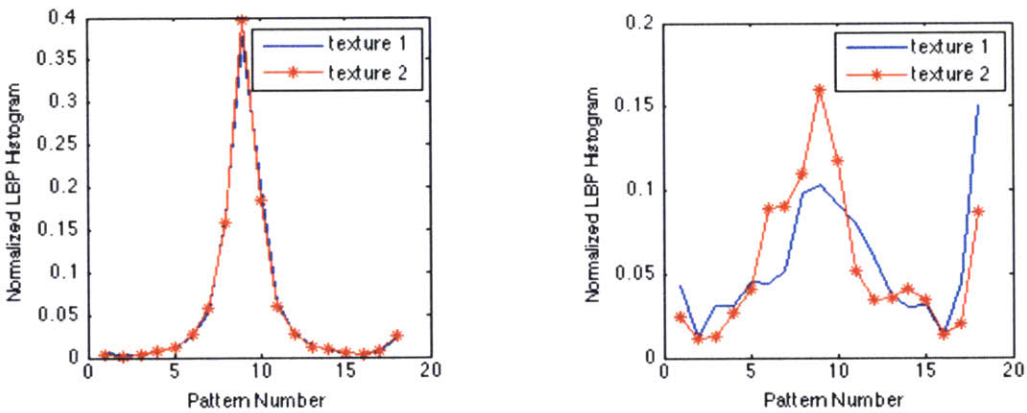
and the superscript *riu2* reflects the rotation invariance "uniform" pattern with *U* not more than 2. In practice, $LBP_{P,R}^{riu2}$ has *P* + 2 distinct output values. The texture feature employed is the histogram of the operator outputs accumulated over a texture sample. $LBP_{P,R}^{riu2}$ is invariant to gray scales and rotation, making it a potentially good candidate for classifying GelSight texture images.

The "uniform" patterns in LBP are indications for structures such as spots, flat areas, and edges of varying positive and negative curvatures. The choice of *P* and *R* affects directly the size of the structures under investigation. Intuitively, the larger the *R* is, the larger the size of the patterns examined; a small *R* corresponds to microstructures and a large *R* macrostructures. As discussed in [9], however, *P* and *R* are closely related and

practically limited by requirements of efficient implementations. Firstly, on one hand, for a given R, we want a large P to reduce the quantization level of the neighborhood circle,



(a) (b)



(c) (d)

Figure 5-3: Illustration of two textures with similar microstructures but very different macrostructures. (a) Texture 1. (b) Texture 2. (c) Histogram of LBP values for $(P,R) = (16,2)$ on the original images of texture 1 and 2. This represents statistics of the microstructures with $R = 2$. (d) Histogram of LBP values for $(P,R) = (16,2)$ on images with reduced dimensions at level 2 of Gaussian pyramid (See Section 5.2.2). Equivalently this represents statistics of the macrostructures with $R = 8$ in the original image.

which is determined by $360°/P$. On the other hand, circular neighborhood for a given R contains a limited number of pixels (e.g., 8 for $R = 1$), which sets an upper limit for P in order to avoid redundancy in calculating the LBP value. A sensible relationship between P and R is that $P = 8R$. Secondly, an efficient implementation with a lookup table of 2^P

elements sets an upper limit to P for real-time applications [9]. For example, with (P,R) = (32,4), the size of the lookup table for $LBP_{P,R}^{riu2}$ can be up to 4 Gigabytes which is quite big and it becomes slow to find a particular match of the LBP value in the lookup table. This may limit the potential application of the algorithm for real-time implementations. For the above two reasons, Ojala et al. [9] only considered (P,R) values of (8,1), (16,2) and (24,3). However, this limits the capabilities of using macrostructures with $R > 3$ as texture features with larger P and R. In fact, many texture images in the real world may contain similar microstructures but different macrostructures. Figure 5-3 shows an example of two visually very different textures that have similar microstructures but very different macrostructures.

To achieve a high classification rate, it is practically desirable to have operators that include statistics of both microstructures and macrostructures as the texture features without increasing the values of P and R. In the next section, we introduce MLBP that takes into consideration both micro- and macrostructures, and achieve a significant increase in performance.

5.2.2 Multi-Scale Local Binary Pattern

The conventional LBP may not deal with macrostructures effectively for $R > 3$. Instead of increasing P and R on the original texture image as Ojala et al. did [9], we can first reduce the dimensions of the images and then apply the same P and R to obtain the LBP features. We employ the Gaussian pyramid [10] approach for size reduction before applying LBP.

Denote the original image as pyramid level 0, reduced pyramid level as level 1 and so on. Effectively, reducing image dimensions by half before applying LBP is approximately equivalent to applying LBP with twice the P and R on the original image in terms of normalized LBP histograms (regardless of and the weights used to generate the Gaussian pyramid), yet with a much faster speed. For instance, the histogram information contained in the LBP values obtained with (P,R) = (16,2) at level 1 is approximately equivalent to that obtained with (P,R) = (32,4) at level 0 after normalizing the LBP histograms. By considering histograms of LBP values at different pyramid levels of the original image, we can take into consideration both micro- and

macrostructures of different sizes. The feature vector used for classification is a weighted concatenation of LBP histograms at different levels or different image scales. We refer this histogram operator as multi-scale local binary pattern, or MLBP for short.

While the multiresolution LBP by Ojala et al. [9] also tries to combine statistics of structures of different sizes, its fundamental limitation is that it does not go beyond R > 3 due to the reasons discussed in the beginning of Section 3 and, therefore, does not take into consideration statistics of macrostructures. The MLBP operator is different from the multiresolution LBP by Ojala et al. [9] in that we find LBP histograms of different pyramid levels with possibly the same P and R for each level, instead of increasing P and R directly on the original image. In this way, it really enables us to consider statistics of structures at different scales. More specifically, we use n different pyramid levels with parameter $a = 0.375$ in Burt and Adelson paper [10] and obtain the histogram of $LBP_{P,R}^{riu2}$ as in Eqn. (8) for each level $i = 0,\ldots, n\text{-}1$, which we denote as $L_{P,R}^{i}$ for simplicity. We then apply weights w_i to $L_{P,R}^{i}$ and concatenate them to form a big feature vector denoted as:

$$MLBP_{P,R}^{n} = concat\left[w_i \cdot L_{P,R}^{i}\right] \tag{10}$$

for $i = 0,\ldots,n\text{-}1$, where n is the number of pyramid levels, and concat[.] is the concatenation operator. Note that different from previous notations where LBP refers to the pattern over a particular neighborhood of pixels, MLBP here is used to denote the concatenated *histograms* of $LBP_{P,R}^{riu2}$ at different pyramid levels.

For each level reduction, the image size is reduced approximately by half in each dimension, and the number of pixels is reduced to approximately 1/4 of the original image. So is the sum of the LBP histogram. We use the simple weights $w_i = 2^i$ to partially compensate for the reduction of the image size while noting that smaller images play less important, but not negligible, roles in texture classification. For example, for n = 4, and $(P,R) = (16,2)$, we have

$$MLBP_{16,2}^{4} = \left[L_{16,2}^{0} \;\; 2L_{16,2}^{1} \;\; 4L_{16,2}^{2} \;\; 8L_{16,2}^{3}\right] \tag{11}$$

It is possible to use a different set of weights, e.g., based on statistics of the texture database, but we do not see significant improvement over the simple scheme above, while adding in a new class might change the weights completely. Hence we do not go into the detail in this work. Also, an extension to the MLBP operator is to use different (P,R) values at different pyramid levels. We do not see added advantages of doing so, and here we use the same (P,R) value across different levels.

5.2.3 Similarity Metric

For histogram features, Arandjelovic et al. [11] recently proposed that the Euclidean distance measure often yields inferior performance compared to using measures such as Hellinger. In this work, we used Hellinger distance as the similarity metric. The Hellinger kernel, or Bhattacharyya coefficient [12], for two $L1$ normalized histograms, $\boldsymbol{x}$ and $\boldsymbol{y}$, is defined as:

$$BC(\boldsymbol{x},\boldsymbol{y}) = \sum_{i=1}^{n} \sqrt{x_i y_i} \tag{12}$$

where $\sum_{i=1}^{n} x_i = 1$ with $x_i \geq 0$, and $\sum_{i=1}^{n} y_i = 1$ with $y_i \geq 0$.

It was shown in [11] that after the following two steps, comparing Euclidean distances of the resulting vectors is equivalent to comparing Hellinger distances of the original vectors: (i) $L1$ normalize the feature vector so that $\sum_{i=1}^{n} x_i = 1$ with $x_i \geq 0$; (ii) square root each element x_i. This is because the Euclidean distance can then be expressed as:

$$d_E(\sqrt{\boldsymbol{x}},\sqrt{\boldsymbol{y}})^2 = \|\sqrt{\boldsymbol{x}} - \sqrt{\boldsymbol{y}}\|_2^2 \tag{13}$$

$$= 2 - 2BC(\boldsymbol{x},\boldsymbol{y})$$

$$= 2H(\boldsymbol{x},\boldsymbol{y})^2 \tag{14}$$

where $H(\boldsymbol{x},\boldsymbol{y}) = \sqrt{1 - BC(\boldsymbol{x},\boldsymbol{y})}$ is the Hellinger distance. In this way, we have the

flexibility to apply many readily available built-in functions in various image processing software such as MATLAB. The smaller the Hellinger distance, the more similar the two histograms or feature vectors are.

5.3 Experiments

Experiments were done on both Outex databases [12] and GelSight images to test the performance of the MLBP operator. The Outex databases contain 2D texture images that are used to compare performance of MLBP with that of other methods. The GelSight images are of real interest for tactile sensing and are used to validate the performance of MLBP. Here we convert GelSight 3D height maps to 2D gray images by using brightness levels to represent the height information. While there is a clear distinction between 2D visual textures such as those in the Outex databases and the 3D surface textures in GelSight, the basic principle of texture classification remains the same.

5.3.1 Experiment on Outex Databases

The Outex database is a publicly available framework for experimental evaluation of texture analysis algorithms [12]. There are a number of test suites available. We are particularly interested in the following two that are most popular for evaluating texture classification algorithms in terms of invariance to gray scales and rotation:

1. **Test suite Outex_TC_00010 (TC10):** There are 24 textures in total, and each texture contains 180 samples at nine rotation angles (0°, 5°, 10°, 15°, 30°, 45°, 60°, 75°, and 90°). Each sample has dimension 128×128 pixels. Figure 5-4 shows the 24 textures at angle 0°. The classifier is trained with the reference textures (20 samples of illuminant "inca" and angle 0° in each class), while the other 160 samples of the same illuminant but the other eight rotation angles in each texture class, are used for testing the classifier. In this suite, there are 480 training samples and 3,840 testing samples in total.

2. **Test suite Outex_TC_00012 (TC12):** The classifier is trained with the reference textures (20 samples of illuminant "inca" and angle 0° in each class) and tested with all samples captured using illuminant "tl84" and "horizon". In each problem, there are 480 training samples and 4,320 testing samples.

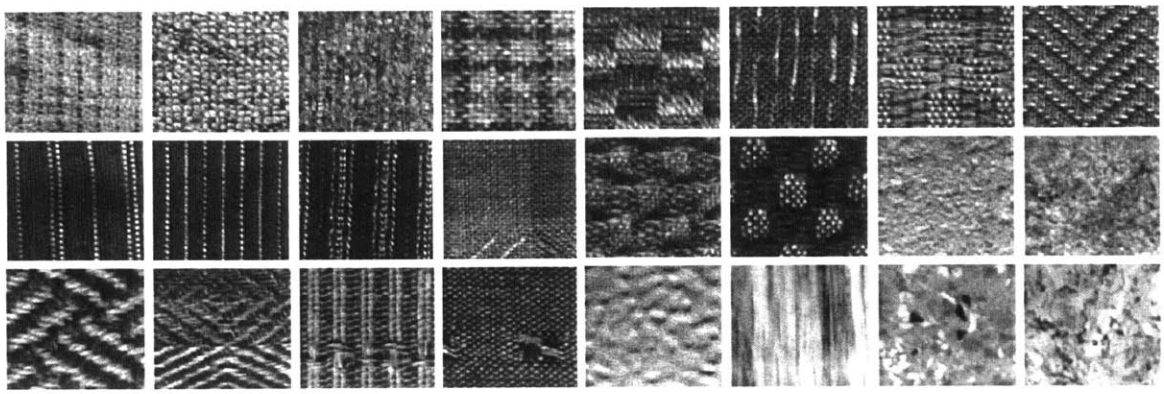


Figure 5-4: 24 textures of dimension 128x128 inOutex_TC_00010 and Outex_TC_00012.

Our goal is to maximize the classification rate, defined as the ratio of the number of correctly classified samples to the total number of samples for classification. First, we find the histogram of MLBP as in Eqn. (11) for each training sample, $L1$ normalize it and square root each element. For a given testing sample, we do the same operation as the training samples, and find its 3 nearest neighbors using Euclidean distance measure among the training samples. With the above operations, equivalently we are using Hellinger distance metric on the histograms of MLBP. Among the 3 nearest neighbors found, if at least two of them belong to the same class, we output that class as the class of the testing sample. If the 3 nearest neighbors belong to 3 different classes, then we output the class of the nearest neighbor as the class of the testing sample.

Nothing prevents us from using different (P,R) pairs at different pyramid levels, but we do not see an added advantage of doing that. Hence we will use the same pair of $(P,R) = (16,2)$ at different pyramid levels for this experiment.

Table 1 shows the correct classification rate using different methods by comparing MLBP with 6 other classical texture classification algorithms: $\mathrm{LBP}_{P,R}^{riu2}$ [9], $\mathrm{LBP}_{P,R}^{riu2}/\mathrm{VAR}_{P,R}$[9], LBP-HF [13], LBPV with global matching ($\mathrm{LBPV}_{P,R}^{u2}\mathrm{GM}_{ES}$) [14], dominant LBP (DLBP) [15], and MR8 [16]. LBP-HF, $\mathrm{LBPV}_{P,R}^{u2}\mathrm{GM}_{ES}$, and DLBP are improved versions of LBP, and their best performances among all (P,R) pairs used by the authors are listed here for comparison. MR8 is the state-of-the-art statistical algorithm for texture classification.

Table 5.1: Correct classification rates (%) of different methods, with the highest rate of each column highlighted in bold.

	TC10 "inca"	TC12 "tl84"	TC12 "horizon"
$LBP_{P,R}^{riu2}$	96.10	88.80	83.40
$LBP_{P,R}^{riu2}/VAR_{P,R}$	97.70	87.30	86.40
LBP-HF	97.97	91.50	87.66
$LBPV_{P,R}^{u2}GM_{ES}$	97.76	95.39	95.57
DLBP	99.10	93.20	90.40
MR8	92.50	90.90	91.10
MLBP	**99.17**	**98.91**	**98.22**

From Table 5.1 it can be seen that for all TC10 and TC12 databases under different illuminations, MLBP achieves the best classification rates among all 7 methods compared and is most invariant to different illuminations. In particular, for the TC10 database, MLBP increases the rate to 99.17% from 96.10% of $\mathrm{LBP}_{\mathrm{P,R}}^{\mathrm{riu2}}$. For the TC12 "tl84" and "horizon" databases, MLBP increases the rate by 10.11% and 14.82% respectively, compared to $\mathrm{LBP}_{\mathrm{P,R}}^{\mathrm{riu2}}$. This shows that the MLBP operator is most invariant to rotations under the same illuminant. Furthermore, when we compare the performance of MLBP for TC10 and TC12 under 3 different illuminants, we see that the classification rate is very stable (98.22% ~ 99.17%) as compared to other methods such as DLBP. This means that MLBP is also invariant to gray scales to a large extent. Nevertheless, when the illumination is not uniform, which is often the case in real-world conditions, all the above methods may not perform well simply due to the fact that shadows or illuminants now become part of the textures. Hence it becomes beneficial for us to use GelSight height images combined with MLBP to classify those textures.

5.3.2 Experiment on GelSight Databases

We obtained 40 classes of GelSight texture images from the GelSight portable device (Figure 2-3(b)) using the techniques described in [2]. Each class consists of 6 texture images at random orientations and with dimensions 480×640 pixels. Each image is then cropped to 4 non-overlapping samples of dimension 200×200 pixels at the center of the

image, with 960 samples in total.

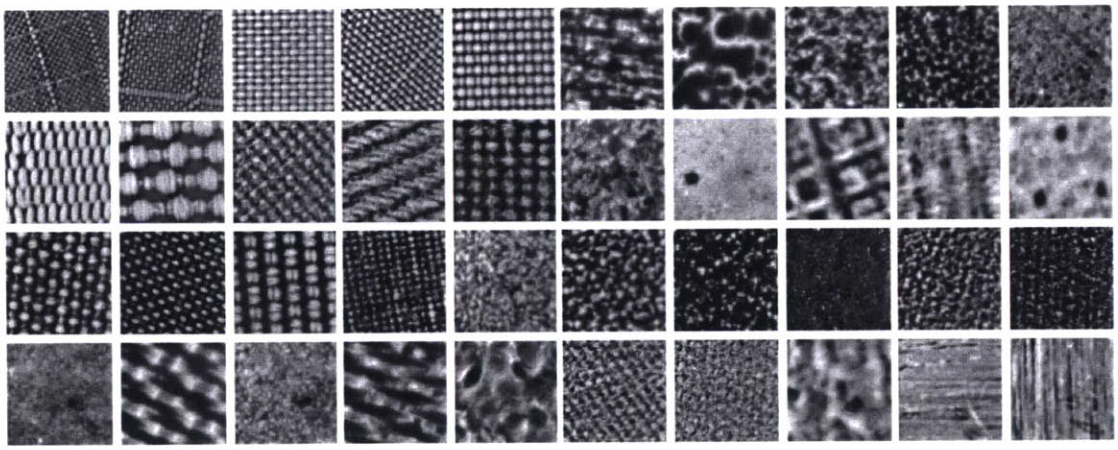


Figure 5-5: GelSight texture database with 40 different texture classes, comprised of 14 fabrics, 13 foams, 3 tissue papers, 7 sandpapers, 1 plastic and 2 wood textures.

Table 5.2: Correct classification rate (%) of MLBP for different numbers of training samples with the highest rate highlighted in bold.

Number of training samples per texture	MLBP $(P,R) = (16,2)$	MLBP $(P,R) = (8,1)$
8	99.22	98.12
12	**99.79**	98.96
16	99.69	99.38

The actual images obtained from GelSight are height maps. We convert them to 2D images with brightness of pixels indicating the height levels: the brighter the pixel in the corresponding 2D image, the larger the height is. Figure 5-5 shows samples of the 40 texture classes. Numbered from left to right and up to down, the surfaces are 14 fabrics, 13 foams, 3 napkins, 7 sandpapers, 1 plastic and 2 woods. Note that the database contains some really similar textures, such as textures 1 and 2, 17 and 19, 15, 16, 18, and 25, etc., which makes the classification task challenging. Among the 24 samples for each texture class, some are used as the training samples and the rest as testing samples. Table 5.2 shows the correct classification rate for different numbers of training and testing samples. Here we use $n = 4$ pyramid levels.

It can be seen that among all the different settings, MLBP with $(P,R) = (16,2)$ and 12 training samples can give the best performance of 99.79% with only one sample

misclassified out of 480 testing samples. As the number of training samples increases, the correct classification rate is expected to increase as well, as there are more samples to be compared with. But the classification speed may become to decrease. In practice, we will find a compromise between the number of training samples used and speed especially when the classification is performed in real time, such as in the case of robotic tactile sensing. Also, we may use different (P,R) pairs for different tasks.

Tactile sensing is an important but challenging area for robotics. With the compliant properties of gel elastomers that mimic human fingers, GelSight is a promising candidate for tactile sensing and material perception. This work focuses on the classification of surface textures, where the texture data is based on height maps attained by touching a surface with a GelSight sensor. We adopted techniques based on local binary patterns (LBP). Conventional LBP and improved versions such as LBP-HF and DLBP mainly look at microstructures of textures and overlook the macrostructures that may be important distinguishing features for different textures. In this Chapter, we presented a novel multi-scale operator, MLBP, that takes into consideration both microstructures and macrostructures for feature extraction. We also adopted the Hellinger distance as a similarity metric. To compare our algorithm with current techniques in the visual texture literature, we used the Outex databases. MLBP performed the best among several classical methods for texture classification. We also built a database of GelSight surface textures, with 40 classes of different materials, and achieved a classification rate as high as 99.79%. Although the database is small, the high classification rate indicates that our system is well suited to the task of recognizing high-resolution surface textures, and may help to deliver a rich form of information for robotics.

Chapter 6

Conclusion

For robots to perform advanced manipulation in a world of unknowns, touch is a critical source of information, and a high-quality tactile sensor is essential. GelSight converts pressure patterns to images that may be handled using image processing or computer vision techniques, making it promising for advanced tactile sensing. In this work, we developed a novel tactile sensor based on GelSight, called the fingertip GelSight sensor, that has ultra high resolution, on the order of tens of microns, high compliance and high sensitivity. We redesigned the previous versions of GelSight hardware and software to make it much more compact, faster and less expensive, and mountable on a robot gripper. We demonstrated its unparalleled capabilities as a new-generation robotic fingertip sensor for versatile manipulation.

Equipped with a fingertip GelSight sensor, the Baxter robot is now able to perform small parts manipulation such as USB insertion with a high success rate, which requires sub-millimeter accuracy and would be difficult with other types of tactile sensors. We also used the fingertip sensor for normal and shear force estimation, as well as slip

detection, which are critical for automated manipulation in the world of unknowns. With such information, the robot is now able to pick up and manipulate a wide range of objects with an adequate amount of force required to keep them stable while not crushing them, including bottles, eggs, potato chips and even leaves. Furthermore, we developed algorithms for material recognition with GelSight, in terms of 3D surface texture classification with a high success rate. With image processing and machine learning on the tactile images obtained by GelSight, the fingertip GelSight sensor opens many possibilities for robotic manipulation based on tactile sensing, which would otherwise be difficult to achieve.

6.1 Limitations

Despite the fact that the current sensor is a lot smaller than previous versions, it is relatively thick (~30mm) and has a small sensing area (~13mmx17mm). The reason is that the webcam has a moderate field of view (90°): increasing the distance between the camera and the gel will increase the sensing area but that also increases the thickness. This issue might be addressed by using a wide-angle lens to compress the thickness while increasing the sensible area and/or using a different design with mirror reflections.

Furthermore, the sensor does not measure vibration or temperature information, but these modalities may be added in future designs.

6.2 Future Work

Besides the future work addressed in the limitations above, future research should also consider the following.

Design finger-shaped sensor. The current shape of the sensor is a cubic box which may be inconvenient for complex manipulations. In future, more compact finger-shaped sensors may be designed. To do so, the illumination system needs to be redesigned, and a smaller camera with a wider field of view may be employed.

3D printing of gels. Currently the gels are made in our lab manually. The quality or properties of each one may vary and hard to be controlled well for consistency. In future, we may use 3D printing to print the transparent gels together with the supporting plate.

Active sensing with GelSight. With a fingertip GelSight sensor, it is possible to perform active sensing to estimate object properties such as softness and roughness, and also to sense shape information such as edges and curvatures. Due to time limit, these aspects were not pursued and will be left for future work.

Bibliography

[1] M. K. Johnson, and E. H. Adelson, “Retrographic sensing for the measurement of surface texture and shape”, Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Florida, USA, 2009.

[2] M. K. Johnson, F. Coley, A. Rajz and E. H. Adelson, "Microgeometry Capture using an Elastomeric Sensor", SIGGRAPH 2011.

[3] R. Li, and E. H. Adelson, "Sensing and Recognizing Surface Textures Using a GelSight Sensor," IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Oregon, USA, 2013.

[4] X. Jia, R. Li, M. A. Srinivasan, and E. H. Adelson, "Lump Detection with a GelSight Sensor," IEEE World Haptics Conference, Daejeon, Korea, 2013.

[5] R. Li, R. Platt Jr, W. Yuan, A. Pas, N. Roscup, M. A. Srinivasan, and E. Adelson, “Localization and Manipulation of Small Parts Using GelSight Tactile Sensing,” in IEEE International Conference on Intelligent Robots and Systems (IROS), 2014, pp. 3988–3993.

[6] P. P. Systems, “Products - RoboTouch,” http://www.pressureprofile.com/products-robotouch.php.

[7] R. Diankov, "Openrave," *http://openrave.org*.

[8] I. Sucan and S. Chitta, "Moveit!" *http://moveit.ros.org*.

[9] T. Ojala, M. Pietikainen and T. Maenpaa, "Multiresolution Gray-Scale and Rotation Invariant Texture Classification with Local Binary Patterns," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 24, no. 7, pp.971–987, 2002.

[10] P. J. Burt and E. H. Adelson, "The Laplacian pyramid as a compact image code," IEEE Trans. Commun., vol. 31, no. 4, pp. 532-540, April 1983.

[11] R. Arandjelović, A. Zisserman. "Three things everyone should know to improve object retrieval," IEEE Conference on Computer Vision and Pattern Recognition, 2012.

[12] T. Ojala, T. Maenpaa, M. Pietikainen, J. Viertola, J. Kyllonen, and S. Huovinen, "Outex – A New Framework for Empirical Evaluation of Texture Analysis Algorithms," *Proc. 16th Int'l Conf. Pattern Recognition*, 2002.

[13] T. Ahonen, J Matas, C. He and M. Pietikainen. Rotation invariant image description with local binary pattern histogram Fourier features. *SCIA Proc.*, pages 61-70, 2009.

[14] Z. Guo, L. Zhang, and D. Zhang. Rotation invariant texture classification using LBP variance (LBPV) with global matching. Pattern Recognition, pages 706-719, 2010.

[15] S. Liao, M. W. K. Law, and A. C. S. Chung, "Dominant Local Binary Patterns for Texture Classification," IEEE Transactions on Image Processing, vol. 18, no. 5, May 2009.

[16] M. Varma and A. Zisserman. A statistical approach to texture classification from single images. *International Journal of Computer Vision: Special Issue on Texture Analysis and Synthesis*, 62(1--2): 61--81, April 2005.

[17] M. Pietikainen, T. Ojala, Z. Xu, Rotation-Invariant texture classification using feature distributions, Pattern Recognition 33 (2000) 43–52.

[18] Fishel, J.A., Loeb, G.E., Bayesian exploration for intelligent identification of textures. *Frontiers in Neurorobotics* 6:4, 2012

[19] R. M. Crowder. Toward robots that can sense texture by touch. Science, 312:1478–1479, 2006.

[20] V. Maheshwari and R. F. Saraf. High-resolution thin-film device to sense texture by touch. Science, 312:1501–1504, 2006.

[21] M. Knibestöl, Å.B. Vallbo, Single unit analysis of mechanoreceptor activity from the human glabrous skin, Acta Physiol. Scand. 80 (1970) 178–195.

[22] R.S. Johansson, J.R. Flanagan, Coding and use of tactile signals from the fingertips in object manipulation tasks, Nat. Rev. Neurosci. 10 (May) (2009) 345–359.

[23] H. Yousef, M. Boukallel, K. Althoefer, Tactile sensing for dexterous in-hand manipulation in robotics – A review, Sens. and Actuators A: Phys. 167 (2011) 171-187.

[24] Y. Xu, Y.-C. Tai, A. Huang, C.-M. Ho, IC-integrated flexible shear-stress sensor skin, J. Microelectromech. Syst. 12 (2003) 740–747.

[25] R.B. Katragadda, Y. Xu, A novel intelligent textile technology based on silicon flexible skins, Sens. Actuators A: Phys. 143 (May) (2008) 169–174.

[26] R.B. Katragadda, Y. Xu, A novel intelligent textile technology based on sil- icon flexible skins, in: Proceedings. 9th IEEE International Symposium on Wearable Computer, 2005, pp. 78–81.

[27] K. Kim, K.R. Lee, W.H. Kim, K. Park, T. Kim, J. Kim, J.J. Pak, Polymer-based flexible tactile sensor up to 32 × 32 arrays integrated with interconnection terminals, Sens. Actuators A: Phys. 156 (December) (2009) 284–291.

[28] J. Engel, J. Chen, C. Liu, Development of polyimide flexible tactile sensor skin, J. Micromech. Microeng. 13 (May) (2003) 359–366.

[29] K. Kim, K.R. Lee, D.S. Lee, N.-K. Cho, W.H. Kim, K.-B. Park, H.-D. Park, Y.K. Kim, Y.-K. Park, J.-H. Kim, A silicon-based flexible tactile sensor for ubiquitous robot companion applications, J. Phys. Conf. Ser. 34 (2006) 399–403.

[30] W.-C. Choi, Polymer micromachined flexible tactile sensor for three-axial loads detection, Trans. Electr. Electron. Mater. 11 (3) (2010) 130–133.

[31] E.-S. Hwang, J.-H. Seo, Y.-J. Kim, A polymer-based flexible tactile sensor for both normal and shear load detections and its application for robotics, J. Microelectromech. Syst. 16 (3) (2007) 556–563.

[32] Y. Huang, M. Sohgawa, K. Yamashita, T. Kanashima, M. Okuyama, M. Noda, H. Noma, Fabrication and normal/shear stress responses of tactile sensors of polymer/Si cantilevers embedded in PDMS and urethane gel elastomers, IEEJ Trans. Sens. Micromachines 128 (2008) 193–197.

[33] M. Sohgawa, Y. Huang, K. Yamashita, T. Kanashima, M. Noda, M. Okuyama, H. Noma, Fabrication and characterization of silicon-polymer beam structures for cantilever-type tactile sensors, in: Proceedings of the International Con- ference on Solid-State Sensors, Actuators and Microsystems (TRANSDUCERS), 2007, pp. 1461–1464.

[34] V.A. Ho, D.V. Dao, S. Sugiyama, S. Hirai, Analysis of sliding of a soft fingertip embedded with a novel micro force/moment sensor: simulation, experiment, and application, in: Proceedings of the IEEE International Conference on Robotics and Automation (ICRA), 2009, pp. 889–894.

[35] K. Noda, K. Hoshino, K. Matsumoto, I. Shimoyama, A shear stress sensor for tactile sensing with the piezoresistive cantilever standing in elastic material, Sens. Actuators A: Phys. 127 (March) (2006) 295–301.

[36] K. Noda, K. Matsumoto, I. Shimoyama, Flexible tactile sensor sheet with liquid filter for shear force detection, in: Proceedings of the IEEE 22nd International Conference on Micro Electro Mechanical Systems (MEMS), Sorrento, Italy, 2009, pp. 785–788.

[37] K. Noda, Y. Hashimoto, Y. Tanaka, I. Shimoyama, MEMS on robot applications, in: Proceedings of the Solid-State Sensors, Actuators and Microsystems Conference (TRANSDUCERS). International, 2009, pp. 2176–2181.

[38] L. Beccai, S. Roccella, L. Ascari, P. Valdastri, A. Sieber, M. Carrozza, P. Dario, Development and experimental analysis of a soft compliant tactile microsen- sor for anthropomorphic artificial hand, IEEE ASME Trans. Mechatron. 13 (2008) 158–168.

[39] L. Ascari, P. Corradi, L. Beccai, C. Laschi, A miniaturized and flexible optoelectronic sensing system for tactile skin, J. Micromech. Microeng. 17 (November) (2007) 2288–2298.

[40] C. Wen, W. Fang, Tuning the sensing range and sensitivity of three axes tactile sensors using the polymer composite membrane, Sens. Actuators A: Phys. 145–146 (2008) 14–22.

[41] C. Liu, Recent developments in polymer MEMS, Adv. Mater. 19 (2007) 3783–3790.

[42] L.C. Tsao, D.R. Chang, W.P. Shih, K.C. Fan, Fabrication and characterization of electro-active polymer for flexible tactile sensing array, Key Eng. Mater. 381–382 (2008) 391–394.

[43] S. Yu, D. Chang, L. Tsao, W. Shih, P. Chang, Porous nylon with electro-active dopants as flexible sensors and actuators, in: Proceedings of the IEEE 21st International Conference on Micro Electro Mechanical Systems (MEMS), Tucson, AZ, USA, 2008, pp. 908–911.

[44] S. Brady, K.T. Lau, W. Megill, G.G. Wallace, D. Diamond, The development and characterisation of conducting polymeric-based sensing device, Synth. Met. 154 (1–3) (2005) 25–28.

[45] J. Wang, H. Sato, C. Xu, M. Taya, Bioinspired design of tactile sensors based on Flemion, J. Appl. Phys. 105 (April) (2009) 083515–83517.

[46] K. Weiss, H. Worn, The working principle of resistive tactile sensor cells, in: Proceedings of the IEEE International Conference on Mechatronics and Automation, Vol. 1, 2005, pp. 471–476.

[47] Z. Del Prete, L. Monteleone, R. Steindler, A novel pressure array sensor based on contact resistance variation: metrological properties, Rev. Sci. Instrum. 72 (February) (2001) 1548–1553.

[48] H. Alirezaei, A. Nagakubo, Y. Kuniyoshi, A highly stretchable tactile distribution sensor for smooth surfaced humanoids, in: Proceedings of the 7th IEEE-RAS International Conference on Humanoid Robots (Humanoids), 2007, pp. 167–173.

[49] H. Alirezaei, A. Nagakubo, Y. Kuniyoshi, A tactile distribution sensor which enables stable measurement under high and dynamic stretch, in: Proceedings of the IEEE Symposium on 3D User Interfaces (3DUI), 2009, pp. 87–93.

[50] Y.-J. Yang, M.-Y. Cheng, W.-Y. Chang, L.-C. Tsao, S.-A. Yang, W.-P. Shih, F.-Y. Chang, S.-H. Chang, K.-C. Fan, An integrated flexible temperature and tactile sensing array using PI-copper films, Sens. Actuators A: Phys. 143 (May) (2008) 143–153.

[51] Y.-J. Yang, M.-Y. Cheng, S.-C. Shih, X.-H. Huang, C.-M. Tsao, F.-Y. Chang, K.-C. Fan, A 32 × 32 temperature and tactile sensing array using PI-copper films, Int. J. Adv. Manuf. Technol. 46 (2010) 945–956.

[52] M.-Y. Cheng, C.-M. Tsao, Y. Lai, Y.-J. Yang, A novel highly-twistable tactile sensing array using extendable spiral electrodes, in: Proceedings of the IEEE 22nd International Conference on Micro Electro Mechanical Systems (MEMS), 2009, pp. 92–95.

[53] M.-Y. Cheng, C.-M. Tsao, Y.-J. Yang, An anthropomorphic robotic skin using highly twistable tactile sensing array, in: Proceedings of the IEEE Conference on Industrial Electronics and Applications (ICIEA), 2010, pp. 650–655.

[54] M. Shimojo, A. Namiki, M. Ishikawa, R. Makino, K. Mabuchi, A tactile sen- sor sheet using pressure conductive rubber with electrical-wires stitched method, IEEE Sens. J. 4 (2004) 589–596.

[55] T. Someya, T. Sekitani, S. Iba, Y. Kato, H. Kawaguchi, T. Sakurai, A large-area, flexible pressure sensor matrix with organic field-effect transistors for artifi- cial skin applications, Proc. Natl. Acad. Sci. U.S.A. 101 (July) (2004) 9966–9970.

[56] N. Wettels, V. Santos, R. Johansson, G. Loeb, Biomimetic tactile sensor array, Adv. Robot. 22 (2008) 829–849.

[57] C.H. Lin, T.W. Erickson, J.A. Fischel, N. Wettels, G.E. Loeb, Signal processing and fabrication of a biomimetic tactile sensor array with thermal, force and microvibration modalities, in: Proceedings of the IEEE International Confer- ence on Robotics and Biomimetics (ROBIO), 2009, pp. 129–134.

[58] 1.8mm LEDs. http://www.electron.com/leds/through-hole-leds/1.8mm-leds/

[59] Fast Poisson Solver. http://web.media.mit.edu/~raskar/photo/code.pdf

[60] W. Yuan, R. Li, M.A. Srinivasan, and E.H. Adelson, Measurement of Shear and Slip with a GelSight tactile sensor, IEEE International Conference on Robotics and Automation (ICRA), 2015.

[61] Friction. http://en.wikipedia.org/wiki/Friction

[62] J. Craig, Introduction to Robotics, Third Edition. Pearson Prentice Hall, 2005.

[63] W. Bluethmann, R. Ambrose, M. Diftler, E. Huber, A. Fagg, M. Rosenstein, R. Platt, R. Grupen, C. Breazeal, A. Brooks, A. Lockerd, R. Peters, O. Jenkins, M. Mataric, and M. Bugajska, "Building an autonomous humanoid tool user," in IEEE Int'l Conf. on Humanoid Robots, 2004.

[64] R. Murray, Z. Li, and S. Sastry, A Mathematical Introduction to Robotic Manipulation. CRC Press, 1994.

[65] B. Yoshimi and P. Allen, "Integrating real-time vision and manipulation," in Proc. of 13th Hawaii Int'l Conf. on System Sciences, vol. 5, January 1997, pp. 178–187.

[66] W. Meeussen, M. Wise, S. Glaser, S. Chitta, C. McGann, P. Mihelich, E. Marder-Eppstein, M. Muja, V. Eruhimov, T. Foote, J. Hsu, R. Rusu, B. Marthi, G. Bradski, K. Konolige, B. Gerkey, and E. Berger, "Autonomous door opening and plugging in with a personal robot," in IEEE Int'l Conf. on Robotics and Automation, 2010.

[67] R. Platt, F. Permenter, and J. Pfeiffer, "Using bayesian filtering to localize flexible materials during manipulation," IEEE Transactions on Robotics, vol. 27, no. 3, 2011.

[68] P. Allen and P. Michelman, "Acquisition and interpretation of 3-d sensor data from touch," IEEE Transactions on Robotics and Automation, vol. 6, no. 4, pp. 397–404, 1990.

[69] Y. Jia and M. Erdmann, "Pose and motion from contact," International Journal of Robotics Research, vol. 18, no. 5, pp. 466–490, 1999.

[70] A. Okamura and M. Cutkosky, "Feature detection for haptic exploration with robotic fingers," International Journal of Robotics Research, vol. 20, no. 12, pp. 925–938, 2001.

[71] S. Chhatpar and M. Branicky, "Localization in robotic assemblies with position uncertainty," in IEEE Int'l Conf. on Intelligent Robots and Systems, 2003, pp. 2534– 2540.

[72] K. Gadeyne and H. Bruyninckx, "Markov techniques for object localization with force-controlled robots," in 10th Int'l Conf. on Advanced Robotics, 2001.

[73] A. Petrovskaya, O. Khatib, S. Thrun, and A. Ng, "Bayesian estimation for autonomous object manipulation based on tactile sensors," in IEEE Int'l Conf. on Robotics and Automation, 2006, pp. 707–714.

[74] C. Corcoran and R. Platt, "Tracking object pose and shape during robot manipulation based on tactile information," in IEEE Int'l Conf. on Robotics and Automation, vol. 2, 2010.

[75] Z. Pezzementi, C. Reyda, and G. Hager, "Object mapping, recognition, and localization from tactile geometry," in IEEE Int'l Conf. on Robotics and Automation, 2011.

[76] D. Hristu, N. Ferrier, and R. Brockett, "The performance of a deformable-membrane tactile sensor: Basic results on geometricallydefined tasks," in IEEE Int'l Conf. on Robotics and Automation, 2000.

[77] N. Wettels, L. Smith, V. Santos, and G. Loeb, "Deformable skin design to enhance response of a biomimetic tactile sensor," in IEEE Int'l Conf. on Biomedical Robotics and Biomechatronics, 2008.

[78] E. Torres-Jara, I. Vasilescu, and R. Coral, "A soft touch: Compliant tactile sensors for sensitive manipulation," SAIL Technical Report MITCSAIL-TR-2006-014, 2006.

[79] R. Dahiya, G. Metta, M. Valle, and G. Sandini, "Tactile sensingfrom humans to humanoids," IEEE Transactions on Robotics, June 2009.

[80] S. Leutenegger, M. Chli, and R. Siegwart, "Brisk: Binary robust invariant scalable keypoints," in IEEE Int'l Conf. on Computer Vision, 2011.

[81] M. Fischler and R. Bolles, "Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography," Communications of the ACM, vol. 24, pp. 381–395, 1981.

[82] R. S. Dahiya, G. Metta, M. Valle, and G. Sandini, "Tactile sensingfrom humans to humanoids," Robotics, IEEE Transactions on, vol. 26, no. 1, pp. 1–20, 2010.

[83] H. Yousef, M. Boukallel, and K. Althoefer, "Tactile sensing for dexterous in-hand manipulation in roboticsa review," Sensors and Actuators A: physical, vol. 167, no. 2, pp. 171–187, 2011.

[84] (2014) Tactarray instrumentation solutions – pressure profile systems. [Online]. Available: http://www.pressureprofile.com/tact-array-sensors

[85] N. Wettels, J. Fishel, Z. Su, C. Lin, and G. Loeb, "Multi-modal synergistic tactile sensing," in Tactile sensing in humanoidsTactile sensors and beyond workshop, 9th IEEE-RAS international conference on humanoid robots, 2009.

[86] R. D. Howe, I. Kao, and M. R. Cutkosky, "The sliding of robot fingers under combined torsion and shear loading," in Robotics and Automation, 1988. Proceedings., 1988 IEEE International Conference on. IEEE, 1988, pp. 103–105.

[87] M. A. Srinivasan, J. Whitehouse, and R. H. LaMotte, "Tactile detection of slip: surface microgeometry and peripheral neural codes," Journal of Neurophysiology, vol. 63, no. 6, pp. 1323–1332, 1990.