TABLE OF CONTENTS

Page

LIST OF TABLES

Page

LIST OF FIGURES

LIST OF ABREVIATIONS

ASR	Automatic Speech Recognition
$BaTiO_3$	Barium Titanate
CCD	Charged Coupled Device
CNN	Convolutional Neural Network
FTIR	Frustrated Total Internal Reflection
FFT	Fast Fourier Transform
GMR	Giant Magnetoresistance
ICA	Independent Component Analysis
IMU	Inertial Measurement Unit
k-NN	k-Nearest Neighbors
LED	Light-Emitting Diode
LLE	Local Linear Embedding
MFB	Mel-Filter Bank
PCA	Principal Component Analysis
PDMS	Polydimethylsiloxane
PSD	Power Spectral Density
SMEs	Small and Medium-sized Enterprises
SVM	Support Vector Machine
Taxel	Tactile pixel

TIR	Total Internal Reflection
t-SNE	t-Distributed Stochastic Neighbor Embedding
WRS	World Robot Summit

LIST OF SYMBOLS AND UNITS OF MEASUREMENTS

$\boldsymbol{\alpha}$ or $\boldsymbol{\zeta}$	A sparse vector
A	Area in squared meter
A_r	Real area of contact in squared meter
β or λ	The non-sparsity penalty factor
C	Capacitance in Farad
C_i	A class with index i
$^\circ$ or deg	Degree
d	Distance in meter
$\boldsymbol{D}$	A dictionary of features
$\boldsymbol{d}_j$	A sparse coding feature
ϵ	Permittivity in Farad per meter
E_e	Effective modulus of elasticity in Pascal
f_t or f_{tan}	Tangential force in Newton
f_n	Normal force in Newton
F	Force in Newton
h_f	Height of a tactile sensor's contact area in meter
M'	Distance-normalized moment in Newton
m_z	Moment about the z-axis in Newton-meter
N_{Basis}	Number of features in a sparse coding dictionary

$N(\mu, \sigma)$	A Gaussian distribution with mean μ and standard deviation σ
ρ	Dielectric fill ratio (in %)
ϕ	Adhesive's loading angle in radian or degree
$\phi()$	A regularization function
P_i	Pressure on taxel i in Pascal
σ	Normal stress in Pascal
$s(t)$	An arbitrary temporal signal
τ	Shear stress in Pascal
θ	Wedge angle in radian or degree
$t_{Hamming}$	Hamming window size in millisecond
$t_{Overlap}$	Overlap between two adjacent Hamming windows in millisecond
w_f	Width of a tactile sensor's contact area in meter
$\boldsymbol{X}$	A complete spectrogram
$\boldsymbol{x}^{(i)}$	The segment i of a spectrogram
Hz	Hertz
kHz	Kilohertz
kg	Kilogram
m	Meter
mm	Millimeter
μm	Micrometer

mV	Millivolt
N	Newton
Nm	Newton-meter (torque)
Pa	Pascal
kPa	Kilopascal
s or sec	Second
ms or msec	Millisecond

INTRODUCTION

0.1 Motivation

Over the past decade, developments in the global economy have driven major changes in the way industries produce their goods. Today, there is increasingly varied demand from consumers, requiring robots and other manufacturing systems to deal with a larger variety of products, which also tend to be produced in smaller quantities than before. Consequently, instead of performing a small set of tasks in a repetitive manner as they tended to do in the past, today's robots are rather asked to execute increasingly complex and varied operations. Therefore, in order to adapt today's robots to modern manufacturing environments, current perception technologies and control algorithms must evolve significantly. In this perspective, tactile perception and dexterity have been identified as critical elements for the future of robotics (Georgia Institute of Technology, 2013; University of California San Diego, 2016). They are currently a bottleneck that hinders robots' ability to manipulate and interact with their environment (University of California San Diego, 2016), and advances are necessary to meet current and future automation needs.

0.1.1 Robots in the *agile* manufacturing paradigm

Different factors explain why consumers demand is changing more quickly today than it was fifty years ago. One factor is the globalization of markets, which provides consumers easy access to a wide variety of products from around the world. This allows consumers to discover new products more easily. The ease and speed with which we can now communicate information from a source to a destination located almost anywhere around the world have also contributed to this phenomenon.

In parallel, consumers have been adopting new technologies at a growing speed for the last 150 years. For example, Fig. 0.1 shows the downward trend in the number of years it has taken for important technologies to be adopted by 25% of the US population after appearing on the market.

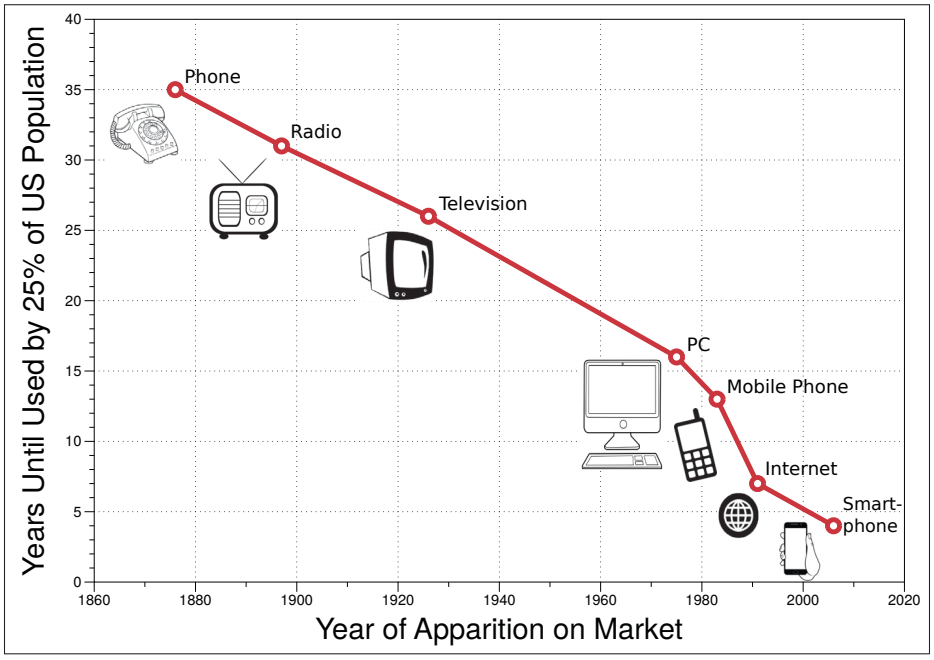


Figure 0.1 Technology adoption rate in the US (data from Ries (2017))

This quick rate of technological adoption, coupled with the fact that consumers now have access to a wider variety of products, has pushed competition between companies to a global level. In the twentieth century, many organizations turned to mass production and quality control to remain competitive—whereas today, a growing number are turning to *lean* and *agile* manufacturing (Devadasan *et al.*, 2012), whose intention is to produce a larger variety of products at lower quantities, with minimal waste. Fig. 0.2 illustrates this paradigm shift.

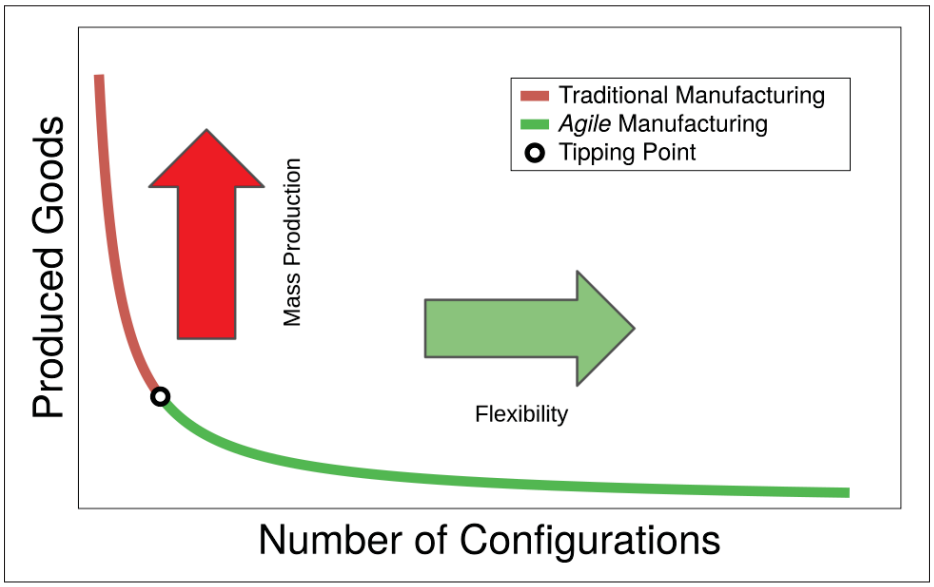


Figure 0.2 Produced goods vs. number of production configurations for agile and traditional manufacturing (image concept inspired by Ries (2017))

In this *agile* manufacturing ideology, traditional automation is not suited to smaller-quantity-but-higher-variety needs, since changes to products or processes generally require costly upgrades and significant downtime. Instead, the ideal manufacturing systems is flexible enough to accommodate a larger number of configurations. For robots to comply with this new context where mass customization is key, both their perception and control algorithms must be improved (University of California San Diego, 2016). Besides, the recent growth in adoption of collaborative robots in factories adds its share of challenges too, since a growing number of robots operate in the same workspace as humans. During the past few decades, research in robotic perception has mostly been focused on providing robots with accurate and reliable artificial vision. Progress in this field has successfully enabled robots to automate a large number of industrial processes. Yet despite significant improvements in artificial vision, many tasks are still difficult or impossible to automate with the current methods of perception employed by industry.

0.1.2 Dexterous manipulation: bridging the gap between humans and robots

The sense of touch helps humans achieve many tasks. Indeed, the somatosensory system (i.e.: touch), with its four main types of mechanoreceptors, gives us sensations that are essential to the execution of a large number of tasks that we often taken for granted. However, people who suffer from hypoesthesia (a reduced sense of touch) have a very hard time doing their everyday activities (Johansson & Flanagan, 2009), highlighting how important this sense really is. These same tasks that are generally easy for humans are frequently difficult to achieve from a robotics perspective. For example, the simple act of lacing a pair of shoes, taking an egg out of its box, or folding a stack of laundry is still quite challenging for robots.

On assembly lines, this gap between robotic and human dexterity severely limits the progression of automation in industries. One of the major limitations affecting industry is the difficulty these robots have in carrying out tasks that require a high level of dexterity, such as assemblies that are highly constrained in force and torque, and particular insertion tasks (Roberge & Duchaine, 2017). This explains why the technical challenge at the 2018 World Robot Summit (WRS) was a complex assembly task that included several insertion steps and aimed at assembling a complete belt drive unit (WRS, 2018): a relatively easy task for human workers, but a very difficult one for robots. Even the simple task of connecting a USB key to a computer port often requires advanced manipulation algorithms to be put in place (Li *et al.*, 2014), whereas humans are able to do this blindfolded using solely their sense of touch.

E-commerce giants also suffer from this lack of advanced dexterous manipulation skills in robots, which limits the automation level in their warehouses. A concrete example of this limitation can be found, for example, in the material handling chains of e-commerce companies such as Amazon, Alibaba, Ebay and many others. These companies have to deal with a wide variety of items and even today, the automation of the packaging step, as well as the handling of items purchased by customers, still remain problematic. These actions are therefore still carried out manually

to a large extent, although robotics is already widely spread inside warehouses (Ackerman, 2019; Ibbotson, 2018). For example, Walmart uses Alphabot robots (Lert, 2020) inside highly automated warehouses to move items purchased by customers, but grabbing the items, aggregating them into orders and placing them into boxes are still performed manually by employees (Ibbotson, 2018). One particular problem with having robots do this task is to implement a control strategy that allows the stable grasping and handling of a particular object without damaging it. Multiple difficulties have prompted a significant proportion of these companies to pay large sums of money to finance research dedicated to the development of technological solutions adapted to these problems. One of the best-known examples in the field is Amazon, which has catalyzed advanced and dexterous manipulation research through the popular Amazon Picking Challenge (Lawrence, 2020). Clearly, industry is motivated to solve the problems brought about by robots' lack of sense of touch.

0.2 Research problems

Given that the sense of touch is at the heart of a robot's interactions with its environment, advances in the field of tactile perception could improve robotic dexterity and allow the further automation of processes that are still operated manually. For this reason, tactile sensors have been proposed for quite some time: they have even appeared as early as the 1970s (Kinoshita *et al.*, 1975). However, despite the added value such sensors could potentially provide by improving robotic manipulation, this technology remains mainly confined to the academic world and research labs: the presence of tactile sensors in industry is still almost non-existent. This section presents the specific research problems that this thesis will address, in hopes of helping remedy this situation.

0.2.1 Leveraging tactile inputs to improve robotic manipulation

One reason why tactile sensors remain unpopular in industry is that there is still no easy way to translate the generally complex raw data generated by tactile sensors into high-level robotic manipulation skills. Indeed, artificial tactile data are often abstract and it is not yet fully understood how they should be used as inputs to control algorithms that could efficiently and reliably solve many of the remaining manipulation problems found in industry. Is it possible to improve robotic dexterity using tactile sensors, and if so, how should they be implemented and used? What tasks can benefit from tactile inputs? Three example tasks that still pose a challenge for robotics, which are nevertheless widespread in different contexts, are discussed below.

- **Handling fragile and deformable objects:** This is still a difficult task in robotics (Shea *et al.*, 2016), and can be considered as both a perception and an actuation problem;
- **Recognizing objects based on touch:** Whereas humans are excellent at recognizing familiar objects using only their sense of touch, such as when digging for spare change in a pocket filled with several other items, robots generally only use vision to locate objects. This often causes problems in occluded and cluttered environments, environments with changing light conditions, or when trying to locate reflective objects;
- **Detecting and classifying important dynamic events:** During manipulation tasks, which involve physical interactions with the environment, many dynamic events occur that can be important to detect and identify in a short period of time. For example, an object slipping out of the gripper's grasp or eventual contact(s) with the environment should be monitored carefully. Many of these dynamic events cannot easily be detected and identified using cameras or typical force-torque sensors found on robots, because these can only provide indirect contact information between the robotic gripper and the environment.

0.2.2 Unravelling tactile sensing modalities

Tactile sensors have significantly evolved since the 1970s and many designs and transduction techniques have now been proposed (Kappassov *et al.*, 2015) in an attempt to replicate some or all of our four types of biological mechanoreceptors. The fact that each of these mechanoreceptors is better at perceiving a specific type of physical interaction adds complexity. For example, given a certain task to accomplish, it is generally difficult to determine how each one of these sensing modalities should intervene. Nevertheless, this has a direct impact on how tactile sensors should be designed and used in robotic tasks.

0.3 Objectives

The main purpose of this thesis is to make steps towards bridging the gap between the seemingly abstract tactile data and their use in advanced manipulation algorithms, in an effort to improve general robotic capabilities. Further:

- A first specific objective is to investigate and better understand if and how tactile data should be encoded and preconditioned, depending on the task, before being used as inputs to control algorithms;
- Thereby, another goal is to study how tactile inputs could be combined with artificial intelligence algorithms to enhance robotic abilities, in order to improve manipulation in the context of the problems mentioned in section 0.2.1. Specifically, tactile-based object recognition, fragile object handling and dynamic event detection are problems that should be investigated;
- Additionally, a specific objective is to study how tactile sensors could be better designed to unlock a finer level of robotic dexterity, which is necessary for the safe handling of fragile and/or deformable objects, among other tasks;

- Another objective is to determine which tactile sensing modality is most valuable for other robotic tasks, especially tactile-based object recognition. Hence, an analysis on the contribution that each tactile sensing modality can add for this specific task should be achieved.

0.4 Contributions

Appendix I presents a list of journal papers, conference papers, and patents that were contributed by the authors during the completion of this PhD thesis.

0.5 Thesis organization and outline

This thesis is organized as follows:

- In chapter 1, we review the literature related to the principal transduction technologies used for tactile sensing. Subsequently, we present the most common data encoding techniques that have been used or that have the potential of being used with tactile sensing. Finally, we discuss some state-of-the-art applications concerning robotic manipulation with tactile sensing;

- In chapter 2, we investigate the problem of effective handling of delicate objects using a typical industrial robotic gripper. Instead of using a complex control strategy to deal with this problem, we rather propose a solution involving gecko-inspired directional adhesives affixed to an industrial robot gripper and tactile sensor. The special adhesive, which has a chevron pattern that was patented at the end of the research project, is able to sustain large shear forces and high torque at low pressure, which is ideal for manipulating fragile objects. We review the design of a tactile sensor to incorporate this new material and introduce a model of adhesion-controlled friction that depends on the contact area measured by the tactile sensors;

- In chapter 3, we address an advanced tactile perception problem that consists of detecting and distinguishing between four types of important tactile events. More precisely, we deal with the problem of identifying: 1) a part slipping out of a gripper's grasp, 2) a part held in a gripper making contact with the environment, 3) the presence of external vibrations, and 4) the absence of any concerning dynamic events (the control group). Whereas chapter 2 is concerned with hardware, we now investigate the software side of the problem: how to properly encode the data. Finally, we demonstrate that using sparse coding as a data-encoding method allows us to highlight the most important features in the input data to efficiently distinguish the four considered classes of events;

- In chapter 4, we study the efficacy of different tactile sensing modalities in the context of a tactile-based object-recognition task. More precisely, we attempt to accurately identify an object from among a set of 50 objects by using various types of data from the tactile sensors: pressure, vibrations (generated by rubbing the gripper/sensor on the object), and proprioception. We show that the modality of perceived object deformation at the fingertips is enough to accurately identify the object. This allows us to propose a quick and grasp-centric exploration method, with minimal operational cost, for recognizing objects without having to rely on vision;

- The final section concludes by summarizing the main contributions of this thesis and by giving several general recommendations.

CHAPTER 1

LITERATURE REVIEW

Any robotic algorithm that uses touch requires a suitable tactile sensing technology. Hence, this chapter begins by introducing and comparing the most common transduction techniques used for tactile sensing in robotics, with a focus on what type(s) of sensing modality they aim to replicate. It then discusses various techniques for encoding the raw data generated by tactile sensors, which is required since the data are often abstract and difficult to directly interpret. These encoding algorithms usually aim at creating compact representations while still preserving and sometimes even promoting important features from the original data. Finally, this chapter presents examples of recent progress in robotic manipulation using tactile sensing.

1.1 Tactile sensing transduction techniques

There are many ways to build a tactile sensor, but they all share one point: they try to replicate, at different levels, at least one function of a biological tactile sensory system. In humans, functions of our biological tactile sensory system are carried out by four distinct types of mechanoreceptors: Merkel's disks, Meissner's corpuscles, Pacinian corpuscles, and Ruffini endings. What we consider our sense of touch is the combined action of these mechanoreceptors, which are individually responsible for a distinct perceptual function (Johnson, 2001). Together, they allow humans to feel light pressure, deep pressure, vibrations, shear, torsion and temperature (Iwamura, 2009). This section presents different transduction technologies that are commonly mentioned in the literature and used to build sensors that can reproduce some of these sensing capabilities. Some of the most famous sensor implementations for each transduction method will be presented and discussed at the same time.

1.1.1 Optical and vision-based sensors

Optical tactile sensors generally consist of light-emitting diodes (LEDs)—a transduction medium which also often acts as the contact interface—and at least one photodetector, such as a charged

coupled device-based (CCD-based) camera or a photodiode. Depending on the implementation, a geometrical change in the transduction medium will change the way light is transmitted, for example by modulating the reflection intensity, altering the received light spectrum, or changing its wavelength, phase, or polarization (Xie *et al.*, 2013). Some optical tactile sensors are based on frustrated total internal reflection (FTIR) (Lavatelli *et al.*, 2019). Another related technique is based on the visual tracking of some known features embedded in the sensor's material while external forces and moments are applied (see Fig. 1.1). These sensors have some advantages over sensors based on other transduction technologies: their spatial resolution is generally high, and they are unaffected by electromagnetic noise. However, optical tactile sensors tend to be bulky, to consume a significant amount of power, and to require more computing power to process their data (Kappassov *et al.*, 2015).

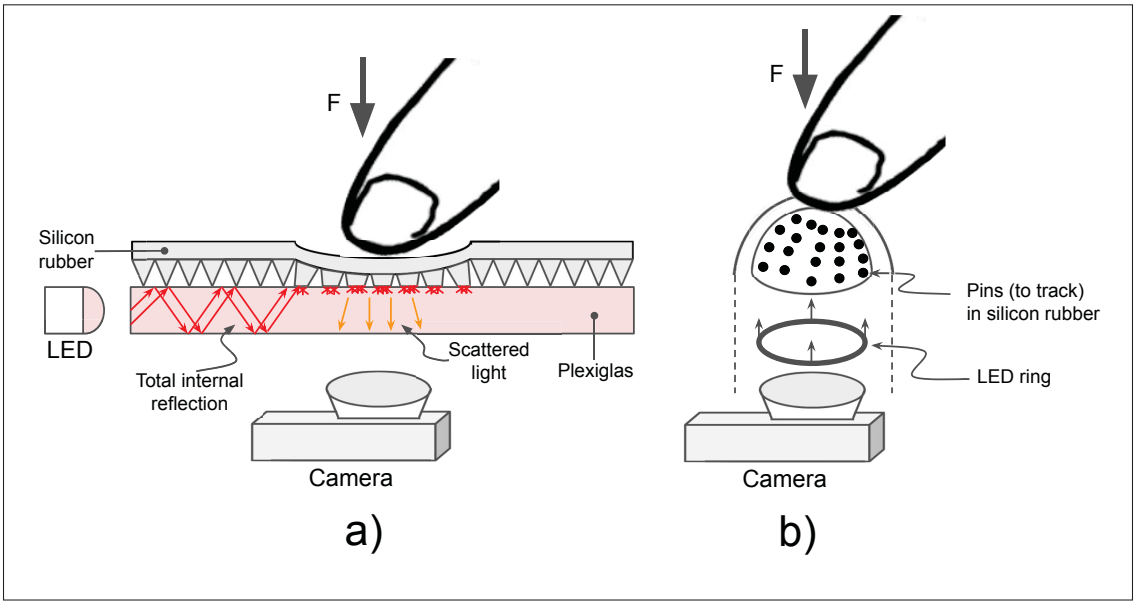


Figure 1.1 Two examples of tactile sensors that use light as a means of transduction: a) optical tactile sensor, in which the contact properties are determined based on FTIR; b) vision-based tactile sensor, in which the contact properties are determined by visually tracking the pins embedded in a silicon material

Early examples of optical tactile sensor implementations were suggested by Begej (1988). The principle was relatively straightforward and consisted of using a camera to measure the frustration of the total internal reflection (TIR) happening on an elastomer interface. While this

specific sensing method is relatively simple to achieve, it still requires a bulky implementation and can mostly only be used to localize and to quantify contact area.

Today, many researchers are choosing vision-based tactile sensors. A good example of such sensors are those from the *TacTip* family (Ward-Cherrier *et al.*, 2018), where a camera is used to track a group of pins embedded within a silicon material with skin-like smoothness. These specific sensors are able to accurately localize contact points with an average error ranging from 0.16 – 0.20 mm. However, the fact that these sensors measure from 85 to 161 mm (from the base to the sensing pad) could make it difficult to integrate them into a robotic gripper, because they could severely limit the stroke of most grippers. Another well-known example of a vision-based sensor is the *GelSight* tactile sensor, which is one of the tactile sensors with the highest spatial resolution (Johnson & Adelson, 2009). It has even been shown that this sensor can detect the difference in height caused by the ink on the surface of a twenty-dollar bill. Although this sensor has been integrated to a robotic gripper (Li *et al.*, 2014), further concerns about its size have pushed researchers to develop a revised and more compact version called *GelSlim* (Donlon *et al.*, 2018). This version uses a mirror to change the camera's field of view, which enabled significant reductions to the sensor's thickness. However, the sensor must now be embedded into a whole finger instead of just a fingertip. Both *GelSight* and *GelSlim* rely on three types of illumination (red, blue, and green) at three different locations and use photometric stereo to convert the images to 2.5D data, which requires more computing power than most transduction technology implementations.

1.1.2 Capacitive sensors

Capacitive tactile sensors normally work by using a ground plane, a smooth dielectric, and a set of sensing electrodes, as depicted in Fig. 1.2. The capacitance between each single electrode and the ground plane is measured using a dedicated chip. From the physics of capacitors (Paul Peter Urone, 2012), we know that

$$C = k\epsilon_0 \frac{A}{d}, \tag{1.1}$$

where C is the capacitance in Farads, k is the dielectric constant of the material, ϵ_0 is the permittivity of free space, A is the overlapping area between the plates in squared meters, and d is the distance separating the plates in meters. As pressure is applied on the sensor, the dielectric will start to be compressed and the individual distances d_i between the electrodes and the ground plane will change accordingly. The capacitance between a specific electrode and the ground plane will change proportionally to the inverse of the distance separating them (it is normally assumed that all other elements in eq. 1.1 will remain constant).

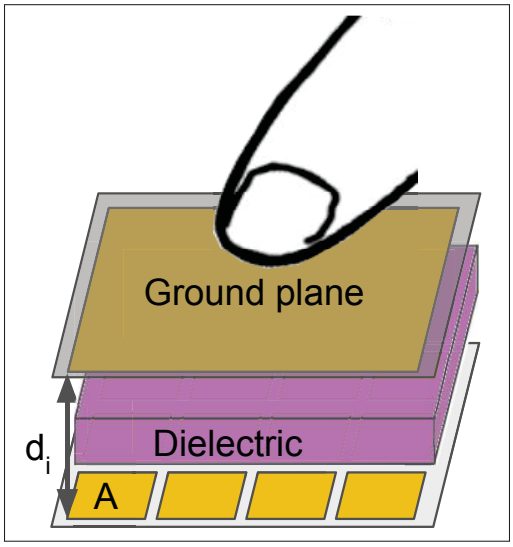


Figure 1.2 Capacitive transduction principle

Following this principle, capacitive transduction has been used frequently in the past to measure normal forces at different points (Charalambides & Bergbreiter, 2013). It has also been successfully implemented by Wu *et al.* (2015) to build a tactile sensor capable of simultaneously measuring forces in three axes and moments about two axes. Rana *et al.* (2016); Le *et al.* (2017) have also used capacitive sensing to measure rapid fluctuations of forces (i.e., vibrations) by amplifying charges flowing in or out of such capacitors using a transimpedance amplifier.

Capacitive tactile sensors have also been used to detect very light contacts (Rana *et al.*, 2016). In fact, the capacitive sensitivity around $F = 0$ N for parallel plates capacitors is expressed by

$$\frac{dC}{dF} \mid_{F=0} = \frac{k\epsilon_0}{\rho d_0 E_e}, \tag{1.2}$$

where ρ is the fill ratio in the dielectric, d_0 is the initial distance between the plates, and E_e is the dielectric's effective modulus of elasticity (Aukes *et al.*, 2014; Wu *et al.*, 2015). Thus, not only is it possible to choose the dielectric material in a way that will increase the sensor's sensitivity to light contacts, but it is also possible to adjust sensitivity by designing the dielectric to achieve specific values for ρ and d_0 in eq. 1.2.

The advantages of capacitive sensing, among all possible transduction technologies, are that such sensors are generally compact, can be highly sensitive, have a relatively good spatial resolution (although normally not as good as optical sensors), are cheap to manufacture, and have a large dynamic range (Tiwana *et al.*, 2012). However, they are unfortunately prone to electromagnetic disturbances. In Gruebele *et al.* (2020), the authors proposed using active shielding to build a stretchable, deformable, and robust tactile sensor. But although this specific capacitive sensor is not affected by electromagnetic noise, the construction of the sensor is more complex than typical implementations.

1.1.3 Piezoresistive sensors and strain gauges

Piezoresistive sensors rely on materials that change their electrical resistance when mechanically deformed, called *piezoresistors*. There are plenty of piezoresistive materials that can be used to make a sensor, as exhaustively reviewed by Stassi *et al.* (2014). The working principle of such sensors basically consists of supplying them with a constant voltage (current) and then reading the current (voltage) while resistance changes with applied pressure on the sensor's piezoresistive interface.

Piezoresistive tactile sensors have been investigated for a long time because they are generally cheap and simple to manufacture. Additionally, they are often thin and conformable, which

allows them to be integrated easily on different devices. For example, Hirai *et al.* (2018) presents a tough, stretchable, and bendable piezoresistive skin to cover robot surfaces. Piezoresistive tactile sensors have also been integrated into different kinds of fabrics (Büscher *et al.*, 2015) and rubbers (Chen *et al.*, 2016b). While some piezoresistive implementations aim to measure normal pressure, a large number are rather focused on measuring shear and/or quantifying skin deformation (Tandon *et al.*, 2015). They normally have a relatively high spatial resolution (comparable or even greater than capacitive sensors) and have a high scanning rate. However, they tend to have a low repeatability and accuracy, to be affected by temperature fluctuations, and to have significant hysteresis (Tiwana *et al.*, 2012).

1.1.4 Piezoelectric sensors

The *piezoelectric effect* designates the generation of an electric charge by certain materials when pressure is applied on them (Fraden, 2003). This piezoelectric ability makes such materials well suited to measuring dynamic variation of forces (i.e., vibrations). In humans, the ability to feel vibrations using the sense of touch allows us to distinguish textures, even those whose spatial periods differ in the scale of hundreds of nanometers (Skedung *et al.*, 2013).

Typical piezoelectric materials include some types of crystals, ceramics, and semi-crystalline polymers. An exhaustive review of piezoelectric materials was made by Ihlefeld (2019). Piezoelectric materials in tactile sensors are useful to measure vibrations. Conversely, such sensors tend not to be efficient at measuring sustained forces applied on their interface. An example of a piezoelectric tactile sensor is given by Dargahi (2000). In this case, the author used this transduction technology to build a three-axis tactile sensor and reported a charge generation in the order of only a few picocoulombs after a force of 1 Newton was applied on the sensor, which illustrates why quantifying static forces would be difficult using this means of transduction. Given its high-frequency response, this technology is generally used for *dynamic tactile sensing* instead. Piezoeletric tactile sensors are also known to have poor spatial resolution and to be affected by temperature fluctuations (Tandon *et al.*, 2015).

1.1.5 Magnetic sensors

Although less common, some tactile sensors use magnetic transduction. These sensors typically rely on a Hall effect sensor or a giant magnetoresistance (GMR) to measure the magnitude of a magnetic field that changes as pressure is applied on the sensor's interface. For example, Alfadhel *et al.* (2016) used a giant magnetoresistance to quantify the magnetic field variation created by the displacement of a magnetic nanocomposite made of polydimethylsiloxane (PDMS) and iron nanowires. Using this configuration, the resulting tactile sensor could measure vertical and shear forces. Another similar example of a tactile sensor capable of measuring normal and tangential forces is presented by Tomo *et al.* (2016), this time using an off-the-shelf Hall effect sensor.

As noted in these works, magnetic tactile sensors are interesting because they are highly sensitive and don't suffer from mechanical hysteresis. On the other hand, they obviously suffer from magnetic interference, and they tend to require complex computations (Tandon *et al.*, 2015).

1.1.6 Barometric sensors

Barometric tactile sensors normally take advantage of a pressure-measuring chip to quantify the pressure variation of a fluid contained close to the sensor's contact interface. For example, RightHand Robotics (2020) used to manufacture a product called TakkTile (2016). These sensors used barometric sensing to measure pressure at different points on the sensor's surface. MEMS barometers enabled the sensor to have a two-gram sensitivity while still being robust (hitting the sensor with a hammer was considered safe) (Tenzer *et al.*, 2014). This sensor has already been integrated to a gripper from RightHand Robotics (2016), specifically by adding TakkTile sensors to each finger and the palm of a *ReFlex Hand* three-finger gripper.

Barometric tactile sensors tend to provide high bandwidth, high sensitivity, and relatively high independence from temperature fluctuations. However, they generally also suffer from poor spatial resolution (Kappassov *et al.*, 2015).

1.1.7 Others

Several other transduction technologies for tactile sensors have been reported in the literature, but they are far less common. Among others, tactile sensors have been reported to be built using ultrasonic transduction (Hutchings *et al.*, 1994), quantum tunnelling (Kulasekera *et al.*, 2013), and structure-borne sound transduction (Kyberd *et al.*, 1998). Many existing sensors are also multi-modal, meaning they combine more than one transduction technology in order to measure more physical variables at the same time, similar to our own somatosensory system. For example, the tactile sensor described in Le *et al.* (2017) uses capacitive sensing to measure pressure at different points on a robotic finger, and an off-the-shelf inertial measurement unit (IMU) to measure vibrations. Many multi-modal tactile sensors include temperature sensing, which is important for humans because it allows them to sharpen their tactile acuity (Stevens, 1982); sensor examples are found in Yang *et al.* (2008); Lee *et al.* (2015).

1.2 Tactile data encoding

Here, the term *encode* refers to the general operation of storing data using representations that will be convenient for the next processing algorithms. For example, during encoding, the raw data can be treated and transformed in an attempt to minimize space or in such a way that only the most important features are highlighted and stored. While data encoding in vision has been widely studied in the context of different kinds of tasks, encoding of tactile data has been far less substantive (Corradi *et al.*, 2014).

Even raw data from human senses must be encoded in order for us to process it. In the case of human vision, although we might believe our eyes provide high-definition vision, we actually only get minimal information. Indeed, a study from Roska & Werblin (2001) has showed that rather than transferring high-definition images to the brain, the optic nerve uses only ten to twelve channels to stream a significantly reduced amount of information. Similarly, when exposed to sound waves, the ear does not send an incredibly rich stream of audio information to the brain. The human cochlea first transforms the sound wave in a manner that is similar to what a

Fourrier Transform would do: it sends the energy (excitation) levels of some frequency bands (i.e., frequency bins) to the brain. It is estimated that even though we can hear frequencies from $20 - 20,000$ Hz, the cochlea sends information for only about $3,000$ frequency bins spanning this full range (Pritchard & Alloway, 1999).

As for the sense of touch, tactile information might also be reduced in size and/or transformed before going through the spinal cord and reaching the brain. For example, Jadhav *et al.* (2009) have studied neural activity of a population of rats. When the rats were whisking across surfaces (i.e., feeling some of the surface properties), they showed that the somatosensory cortex was encoding that information in a probabilistic sparse[1] manner. This provided some clues that there might be a similar data reduction and transformation going on for human tactile data before these data are stored in our neocortex. Consequently, this section lists the most commonly found encoding algorithm in the literature used to encode tactile data.

1.2.1 Principal component analysis (PCA)

Principal component analysis (PCA) is a widespread technique that is used when the data has high dimensionality. Indeed, an important motivation for this approach consists in projecting data in a space of lower dimensions (Smith, 2002). The projection of high-dimensional data to the reduced dimensional space is done by finding axes that maximize the variability of the data. To achieve this, the eigenvalues of the covariance matrix of the original data are first computed. For example, if $\boldsymbol{X}$ is the set of data for which the principal components must be calculated, the covariance matrix $\boldsymbol{C}_X$ is

$$\boldsymbol{C}_X = \left(\boldsymbol{X} - \mu\right)\left(\boldsymbol{X} - \mu\right)^T. \tag{1.3}$$

This matrix is then decomposed to obtain the left eigenvalues (λ) and the left eigenvectors ($\boldsymbol{v}$):

$$\boldsymbol{v}\boldsymbol{C}_X = \lambda \boldsymbol{C}_X. \tag{1.4}$$

[1] Here, *sparse* means that only a few neurons are activated, while most neurons have no significant activity.

By sorting the obtained eigenvalues in descending order $\lambda_1 \geq \lambda_2 \geq \lambda_3 ... \geq \lambda_n$, the corresponding eigenvectors are the components of the data (the new axes), presented in a decreasing order of importance. Knowing the main components of the data, it is subsequently possible to choose a certain number among the most important ones to project the data.

PCA is a dimensionality reduction technique frequently used with tactile data. For example, Edwards *et al.* (2008) rubbed a tactile sensor $1,000$ times across different textures and used PCA to find the 500 most important components in the recorded dataset. This new constructed space was subsequently used to study how well the textures could be distinguished from each other. Aquilina *et al.* (2018) also showed that PCA could enable visualization of similarities in high-dimensional data from four different tactile sensors when they poked different types of cylinders at different locations.

In summary, PCA provides an approximation of some high-dimensional data by projecting them into a new lower-dimensional space. Although doing so generates a certain approximation error, it can also significantly simplify the problem of learning for automatic classification tasks (Harrington, 2012).

1.2.2 Independent component analysis (ICA)

Hérault & Ans (1984) were the first to propose independent component analysis (ICA), a technique that is often associated with the typical problem of source separation and differentiation. A classic example of such a challenging context is the "cocktail party problem," in which several people speak at the same time, generating several superimposed audio frames. The goal is to discern the speech of every person using a specific number of microphones (see Bronkhorst (2000) for an exact and complete description of the problem).

The concept behind this technique differs from PCA. Indeed, the goal here is not to project the data into a reduced space while attempting to preserve as much information as possible, but rather to represent the data $\boldsymbol{x}$ in a linear combination of independent components (Hyvarinen & Oja, 2000). For example, let's suppose there are p tactile sensors that measure signals composed of a

linear combination from n sources. Then, the model assumed by the independent component analysis technique is (Poczos & Tishirani, 2016):

$$\begin{aligned} x_1(t) &= a_{11}s_1(t) + a_{12}s_2(t) + ... + a_{1n}s_n(t) \\ x_2(t) &= a_{21}s_1(t) + a_{22}s_2(t) + ... + a_{2n}s_n(t) \\ &\vdots \qquad\quad \vdots \qquad\quad \vdots \qquad\quad \vdots \\ x_p(t) &= a_{p1}s_1(t) + a_{p2}s_2(t) + ... + a_{pn}s_n(t), \end{aligned} \tag{1.5}$$

with $p \geq n$, and where $x_1(t)...x_p(t)$ are the measurements made by the p tactile sensors at time t, and where $s_1(t)...s_n(t)$ are the superimposed signals from n sources, and $a_{11}...a_{pn}$ are the coefficients that weight the contribution of each signal for a given measurement. It is possible to rewrite eq. 1.5 in matrix form:

$$\boldsymbol{x} = \boldsymbol{A}\boldsymbol{s}. \tag{1.6}$$

The goal is therefore not only to find the coefficients $a_{11}...a_{pn}$, but also to identify each of the sources $s_1(t)...s_n(t)$. (It is considered that the transformation imposed by the matrix $\boldsymbol{A}$ is linear and the measurements $\boldsymbol{x}$ are used to carry out their identification.) To achieve this, different algorithms exist; see Klemm *et al.* (2009) for a detailed list of the main ones. Most of these aim to maximize the decoupling of the $a_{11}...a_{pn}$ coefficients in $\boldsymbol{A}$, in an effort to minimize the information mutually shared by the sources.

Given that $\boldsymbol{A}$ is identified using a specific algorithm, it is possible to encode the tactile sensors' data using the source signals by solving

$$\boldsymbol{s}(t) \approx \boldsymbol{A}^{-1}\boldsymbol{x}(t). \tag{1.7}$$

Similarly to the aforementioned cocktail party problem, encoding tactile data using ICA could allow one to obtain a more logical representation of the signals that are being measured, especially when placed in the context of signal superimposition. For example, Lee *et al.* (2011) dealt with the problem of mixed tactile information from unknown multiple sources. More precisely,

they used tactile sensors mounted on a real robot that made simultaneous contact with several objects. Using ICA, the authors showed they were able to separate the tactile data according to its source (i.e., the object from which it came). This also allowed them to accurately gather tactile information about each specific object.

1.2.3 Local linear embedding (LLE)

Like PCA, local linear embedding (LLE) aims at reducing the dimensionality of a dataset, but using a different criterion. Roweis & Saul (2000) were the first to elaborate this method. The principle consists in projecting data with high dimensions into a space of lower dimensions, while preserving their local neighborhood properties.

The algorithm for doing so starts by expressing each piece of data x_i using the nearest K neighbors as bases, while minimizing the reconstruction error. When all the data ($\boldsymbol{X}$) of the training set has finally been represented as a weighted sum of the nearest K neighbors, the algorithm then proceeds to the next step that aims to reduce the dimension. In this second step the concept is based on the search for vectors that use the linear combinations found in the first step and that allow the data to be reconstructed with a representation error not exceeding a certain threshold. This technique is recognized as being perceptibly more effective than PCA in representing the underlying structure of the data (Shalizi, 2009).

Although it is hard to find examples in the literature where LLE has been used specifically with tactile data, it is easy to find applications of LLE in robotics. For example, Teng *et al.* (2005) created a hand gesture recognition system based on local linear embedding. The authors created features from webcam images using LLE, and their algorithm was able to recognize 30 hand gestures with a classification success rate average of 90%. A similar technique could potentially be used to create features from tactile sensing data.

1.2.4 K-means clustering

K-means clustering is a popular algorithm for which several variants have been proposed. The algorithm was first introduced by Hartigan (1975), although the idea dates back to the early 1960s. The concept, as well as the algorithm, have the merit of being simple, which explains why K-means clustering has been used in so many practical fields. Given a predefined number of clusters K (a number which is manually set by the programmer) and input data $x_i \ \forall i \in \{1, ..., n\}$ where n is the number of data and $x_i \in \mathbb{R}^d$, the goal is to find the coordinates of K clusters, in $\mathbb{R}^d$, which best represent the data. The complete algorithm is described by Hartigan & Wong (1979).

Although this algorithm does not compress the data, it does allow them to be organized by storing data belonging to the same cluster together. K-means clustering can also be used as a classification algorithm for new data, because it can find the cluster they belong to. For example, Luo *et al.* (2014) generated a dataset containing 2,500 tactile data obtained from interacting with 10 different objects. The authors used K-means clustering to create 50 clusters that they subsequently used as features to automatically determine which one of the 10 objects was touched. Repeatedly touching an object and finding the clusters to which the tactile data belonged allowed the object to be correctly identified 91.2% of the time, on average.

1.2.5 Spectral clustering

As mentioned in section 1.2.4, there is a considerable number of variants on the K-means clustering algorithm. Although it would be useless to make an exhaustive description of each of them, the spectral clustering technique is nevertheless important to mention because it is a popular approach used for certain artificial intelligence problems.

If the classical K-Means clustering algorithm can be used to respond to several geometric discrimination problems relatively efficiently, there are still some specific geometric problems that are problematic. A striking example of the classic K-means clustering algorithm's limitation is the case where two distinct classes of data are respectively represented as two distinct spirals intertwined with each other. It is possible to illustrate this problem easily for data in $\mathbb{R}^2$: Fig. 1.3

shows such a layout of the data. In such a case, the classical K-means clustering algorithm will be ineffective in discriminating the two categories of data.

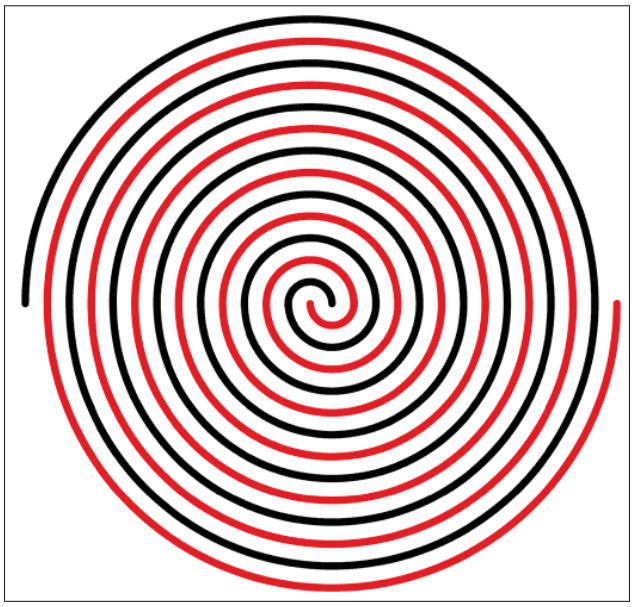


Figure 1.3 Two intertwined spirals representing the arrangement of data from two separate classes in $\mathbb{R}^2$

Spectral clustering, explained in detail in books such as those of Alpert (1995); Ng *et al.* (2001), makes it possible to overcome these difficulties. The approach consists of translating the dataset into a similarity matrix $\boldsymbol{A}$. To do this, several algorithms are available, but the one that seems most popular in the literature is the k-NN algorithm (k-Nearest Neighbors). Once the similarity matrix is obtained, the eigenvalues of the Laplacian ($\boldsymbol{L}_A$) of this matrix are computed. These eigenvalues are sorted in ascending order and the associated first k eigenvectors are used to build a feature vector for each data. Finally, k-means is run to separate the data into k classes. Chebotar *et al.* (2014) studied how to determine the position and orientation of several objects using tactile inputs. They used spectral clustering to reduce the dimensionality of the input data and showed that tactile feedback learning for this task could be achieved more efficiently in this reduced space.

1.2.6 Sparse coding

It was Olshausen & Field (1996) who originally proposed the concept of sparse coding. This technique is based on learning a dictionary of bases (*features*) that are the most important for reconstructing the original data. The mathematical formulation of the dictionary learning problem is given by the following double-optimization problem:

$$\min_{\boldsymbol{D},\boldsymbol{\alpha}} \sum_{i=1}^{m} \left(\left\| \boldsymbol{x}^{(i)} - \sum_{j=1}^{n} \boldsymbol{d}_j \alpha_j^{(i)} \right\|_2^2 + \lambda \sum_{j=1}^{n} \phi\left(\alpha_j^{(i)}\right) \right), \tag{1.8}$$

where $\boldsymbol{D} := [\boldsymbol{d}_1, ..., \boldsymbol{d}_n]$ is the dictionary of features, $\boldsymbol{\alpha}^{(i)}$ is a sparse vector which encodes the input data $x^{(i)}$ as a linear combination of the bases contained in the dictionary, and $\phi(\boldsymbol{\alpha}^{(i)})$ is the regularization function. This latter regularizes the optimization problem by penalizing the use of too many features: without it, the vector $\boldsymbol{\alpha}$ would generally no longer be sparse and would rather point to a large number of bases from the dictionary to reconstruct the input data. Once a dictionary of features has been learned, new data can be encoded into sparse vectors using the solution of the following optimization problem:

$$\min_{\boldsymbol{\alpha}} \left\| \boldsymbol{x}^{(i)} - \sum_{j=1}^{n} \boldsymbol{D} \alpha_j^{(i)} \right\|_2^2 + \lambda \sum_{j=1}^{n} \phi\left(\alpha_j^{(i)}\right). \tag{1.9}$$

Lee *et al.* (2007) showed a way to efficiently solve eq. 1.8 and 1.9. With their proposed method, sparse coding can be implemented and achieved very rapidly on a typical computer. This algorithm could be used as a first step to encode tactile data for further processing. However, very few examples of sparsely encoded tactile data can be found in the literature. One example will be discussed in chapter 3. Another example is the one provided by Liu *et al.* (2016), who used sparse coding to encode tactile data and to find representations that could be used to identify some palpated objects.

By contrast, sparse coding is widely used in other practical fields. Raina *et al.* (2007) used sparse coding to extract high-level features of many different data types (including images, handwritten characters, webpages, and songs) and used the sparse representations to classify the data according to some labels of interest. Since analyzing vibrations acquired by tactile sensors is similar to analyzing sound waves (which are also vibrations, i.e., air vibrations), work on speech recognition and audio classification is also of interest. In this context, Henaff *et al.* (2011) have used sparse representation of songs to automatically classify their genres. Another more advanced technique for sparse coding is the one proposed by Li *et al.* (2013). Here the authors used sparse coding algorithms to build not one, but two dictionaries of high-level bases: a first one in a noisy environment and a second one in a clean context. Using both dictionaries, they achieved in-car speech recognition with a good success rate.

1.2.7 Other encoding techniques

Other encoding techniques could also be used to represent tactile data, but are less commonly-reported in the literature. Additionally, when the tactile data is in the form of a static pressure map, they can often be encoded in the same way images are encoded (Corradi *et al.*, 2014). Some authors (Bartolozzi *et al.*, 2017; Caviglia *et al.*, 2014) have also proposed using an event-driven approach to compress tactile data for robots containing a large number of tactile sensors. The principle is bioinspired and consists of continuously acquiring and measuring tactile data but only transmitting them to a computer for treatment once a significant change in their values is detected.

1.3 Tactile sensing applications in robotics

Tactile sensors can be used in a wide variety of applications; what follows is just a subset of those that are most relevant to this thesis.

1.3.1 Slippage detection and other dynamic event detection

Slippage detection has garnered interest from many researchers, as it is one of the dynamic events most crucial to predicting an upcoming manipulation failure. Indeed, often just before a robot unexpectedly drops an object, it will experience slippage of the handled object with reference to the gripper. Most of the techniques found in the literature to detect slippage are based on handcrafted features. For example, Holweg *et al.* (1996) used two different approaches to detect slippage. The first technique was based on a frequency analysis of the force center displacements on the sensor, whereas the second technique studied the fluctuation of the measured normal force. Melchiorri (2000) also studied slippage detection using a criterion based on a frictional model. In their work, they considered the detection of both linear and rotational slippage using tactile sensing. Goeger *et al.* (2009) proposed the elaboration of a tactile slip sensor that they installed on an anthropomorphic robotic hand. The proposed technique to detect slippage detection method relies on principal component analysis of the frequency spectrum and a handcrafted criterion. Heyneman & Cutkosky (2012) and Heyneman & Cutkosky (2013) worked on distinguishing slippage between robotic fingers and a grasped object, and slippage between a grasped object and an external surface using handcrafted features. While the first type of event (object-gripper slippage) can lead to an object being unexpectedly dropped, the second type (object-world slippage) might not be as alarming and might even be desirable in some contexts. For that reason, they proposed techniques to distinguish the two scenarios based on the use of tactile sensing.

Also found in the literature are machine-learning approaches to the problem of slippage. For example, Shirafuji & Hosoda (2011) used a tactile sensor composed of two sub-sensors: one that uses a piezoelectric component to acquire the vibrations and a strain gauge to acquire some pressure readings. The method they studied used past experiences of slippage to control the force applied by a robotic gripper on an object.

1.3.2 Object recognition, classification and grasping

Tactile sensors can be used alone or in combination with vision to recognize objects, or families of objects, in order to improve a robot's general manipulation abilities. Hyttinen *et al.* (2015) use tactile sensors only to classify the shape of an object and to plan its grasp accordingly. They used a tactile-based model and controlled the gripper's pose depending on the determined class of objects' tactile signatures. Guler *et al.* (2014) used vision and touch, combined and separately, to classify objects prior to robotic grasping. They concluded that even though vision and touch can be used separately for classification with some accuracy, combining both increases the overall classification efficiency.

Vision alone is sometimes only capable of rough localization of the object to be grasped. Consequently, researchers have been investigating how tactile sensors can be used to complement vision in such cases. For example, Dang & Allen (2014) address the problem of grasping with pose uncertainty. To deal with this particular situation, they have shown that one can rely on tactile information to improve the gripper's position with reference to the object to grasp. Vision is also generally of little help when it comes to measuring the contact properties between an object and a robotic gripper. To compensate, Romano *et al.* (2011) used real-time pressure arrays from tactile sensors to control a jaw gripper while grasping and handling an unknown part, and showed that tactile feedback can successfully complement vision in such a context. Yussof *et al.* (2009) also present a tactile sensor that can detect both normal forces and shear in real time. They use this information to improve the robotic grasping, handling, and release of some objects without using vision at all.

1.3.3 Learning how to grasp and handle

Whereas the aforementioned tactile sensing applications aimed at solving some specific manipulation-related problems, they did not provide a framework of artificial intelligence that would allow the robot to learn some aspects of manipulation by itself. In this section, we

explore some papers that intend to improve certain aspects of manipulation by implementing some grasp-and-handle learning algorithms.

One well-known work in this regard is by Google employees Levine *et al.* (2016), who publicized the results of their large-scale data collection (approximately 800, 000 grasp attempts) involving typical pick-and-place tasks. One of the main contributions of this work is certainly the large amount of data they collected, along with the framework with which they were able to gather it. But the principal contribution of this paper is the fact that it is among the first examples of an end-to-end[2] approach to improve robotic manipulation. Indeed, in this paper, the authors used deep learning algorithms (namely, a deep convolutional neural network) to learn, from the sensors' input, what command needs to be sent to the output (the robot's actuators). While this is a recent and very exciting work, the 14 almost-identical robotic manipulation setups they used only relied on monocular vision as sensory input. No such approach has been tested yet with more humanized sensors, such as tactile sensors or stereoscopic vision.

Other works have also tried using learning algorithms to improve robotic manipulation abilities. Bekiroglu *et al.* (2010) instead used hidden Markov models to offline-learn grasp stability using only tactile data. Li *et al.* (2014) used tactile data from previous grasp attempts to adapt the impedance of a three-finger gripper. Each time a new object is grasped, its data is also used to populate the recorded database, and so on. Another well-known piece of research on robotic manipulation is the work by OpenAI *et al.* (2018), in which the authors used several cameras and the Shadow dexterous hand[3] to train a model-free reinforcement learning agent to manipulate a cube. Over time, the agent successfully learned to control the hand in a natural manner and in such a way that it could reorient the cube in different predetermined ways. It is noteworthy that the authors did not use any tactile inputs, and their approach required the use of no fewer than 16 tracking cameras.

---

[2] Here, the term *end-to-end* means that the intelligent algorithms take the sensors' input to directly generate a command to the output actuators, without having to rely on any other software components.

[3] https://www.shadowrobot.com/products/dexterous-hand/

CHAPTER 2

IMPROVING INDUSTRIAL GRIPPERS WITH ADHESION-CONTROLLED FRICTION

Jean-Philippe Roberge[1], Wilson Ruotolo[2], Vincent Duchaine[1], Mark Cutkosky[2]

[1] Systems Engineering Department, École de Technologie Supérieure,
1100 Notre-Dame Ouest, Montréal, Québec, Canada H3C 1K3
[2] Mechanical Engineering Department, Stanford University,
450 Serra Mall, Stanford, California, United States 94305

Paper published in the IEEE Robotics and Automation Letters (RA-L), April 2018.

2.1 Abstract

Effective handling of delicate objects remains a challenging problem in manufacturing. Instead of using a specialized gripper or control scheme we present a solution involving gecko-inspired directional adhesives affixed to an industrial robot gripper and tactile sensor. The adhesives sustain large shear forces with very low pressure. They also release objects without residual adhesion when the grip is relaxed. It is desirable to predict the maximum forces and moments the gripper can exert without slipping. For this purpose the tactile sensor provides an estimate of the area of contact and a force/torque sensor measures the overall force and moment. To resist forces and moments in multiple directions, it is best if the directional adhesives do not all have a single orientation. A chevron pattern strikes a good balance between performance and ease of fabrication.

2.2 Introduction

Achieving a stable grasp while avoiding the application of excessive gripping force is an important objective for many robotic manipulation tasks. The handling of fragile items, for example, is a case where it is crucial for the robot to minimize forces applied to the grasped object to prevent damage. Deformable objects are another class of items where grasping without squeezing is desirable to preserve characteristics such as shape or surface properties.

Humans are highly skilled at performing gentle manipulation, for which we rely on our hands' dense array of mechanoreceptors. We quickly adjust our gripping force in response to our estimate of the object's weight and frictional characteristics formed upon initial contact (Johansson & Westling, 1984; Westling & Johansson, 1984). In reaction to these stimuli, we apply only a minimum force plus a modest safety margin to perform basic manipulations without slippage. This strategy allows us to interact efficiently with fragile and deformable objects. By comparison, most robots have difficulty predicting and maintaining the minimum force required to hold delicate objects.

In this paper we integrate directional, gecko-inspired adhesives into the jaws of a commercial gripper, enabling it to hold very delicate objects and to resist large moments with a small gripping area (fig. 2.1). The adhesives are mounted to the outer surface of a tactile sensor, which provides an estimate of both the area of contact and the gripping force. As we show in later sections, both pieces of information are important for predicting the maximum handling forces and moments that the grip can sustain without slipping. A force/torque sensor at the robot wrist measures the overall external force and moment.

We start with a brief review of related work on which this work builds. Section 2.4 presents the theory behind the design of the gripper's fingertips. The models are confirmed with experiments, initially to characterize the performance of different adhesive orientations and then to evaluate the performance of the entire gripper as mounted on an industrial robot. We discuss the results, present conclusions, and give recommendations for future work.

2.3 Related work

Many approaches have been explored to improve the ability of robots to handle delicate objects. A large portion of the proposed solutions are control-based, and recently, many have relied on tactile sensors to obtain useful information about gripper-object contact properties (Nakamura & Shinoda, 2001; Shinoda *et al.*, 2000; Maria *et al.*, 2015; Cavallo *et al.*, 2014; Deng *et al.*, 2017; Su *et al.*, 2015; Kawamura *et al.*, 2013).

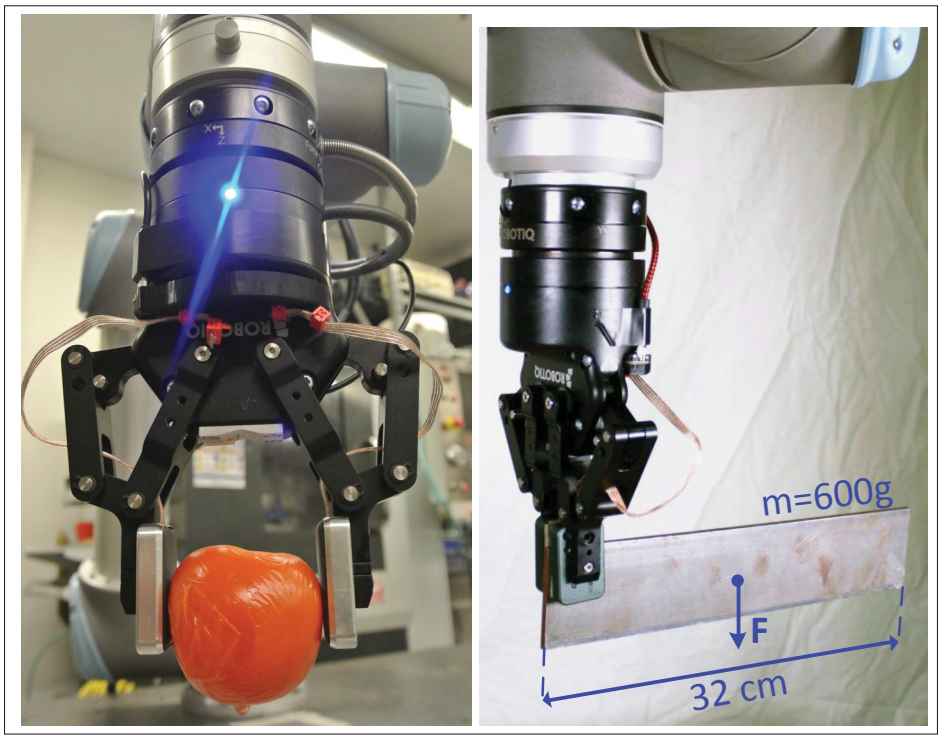


Figure 2.1 An industrial robot, using an arrangement of gecko-inspired adhesives on the gripper pads, can grip and manipulate a rotten tomato, it can also exert considerable torque using only 3/4 of its contact surface (Both applications are presented in the accompanying video, available at http://ieeexplore.ieee.org)

Some authors have proposed preventing slippage by precomputing the required grasping force based on some information that is known *a priori* about the object. In this vein, Nakamura & Shinoda (2001); Shinoda *et al.* (2000) proposed using tactile sensors to rapidly estimate friction coefficients to determine the minimal normal force needed for a successful grasp. Partially based on this concept, Maria *et al.* (2015) proposed using a "tactile exploration" phase to rapidly estimate object-gripper friction using measured tangential forces during initial contacts with the object. Although this method has achieved promising results, it also requires a preliminary motion that takes time and still involves an initial, uninformed force application, which could be problematic for especially delicate objects.

Since slippage occurs in two phases, namely incipient slip and gross sliding (Tremblay & Cutkosky, 1993), other approaches rely on the active detection of either one or both of these phases to re-adjust the exerted grasping force. This concept is widespread in the literature, where different kinds of tactile sensors have been used in conjunction with control strategies to tune the grasping force (Maria *et al.*, 2015; Cavallo *et al.*, 2014; Deng *et al.*, 2017; Su *et al.*, 2015; Kawamura *et al.*, 2013). Many of these works show auspicious results; however, their reliance on complex control strategies and fast response complicates the real-time implementation of such solutions in an industrial setting.

In parallel, numerous novel conformable and adhesive gripping solutions have been proposed (Hawkes *et al.*, 2015; Song *et al.*, 2017; Amend *et al.*, 2012; Tincani *et al.*, 2012, 2013). Among these, Tincani *et al.* (2012, 2013) proposed using *active surfaces* inside a gripper to simulate different levels of friction and apply tangential thrust to a grasped item. With their advantages, however, such strategies also introduce moving parts and complexity to the hand, increasing cost and weight and potentially reducing reliability.

As a consequence, despite these advances, most industrial robots still use either parallel-jaw grippers or suction cups, often with simple control algorithms, as they have done for over 30 years. The solution proposed here can be achieved by retrofitting existing industrial robot grippers and uses gecko-inspired adhesives that allow handling of delicate objects with very low grasp forces. Reviews of gecko-inspired adhesives can be found in Brodoceanu *et al.* (2016); Eisenhaure & D (2014). Among the many possible solutions, we desire a material that is directional and highly controllable, meaning that the magnitude of adhesion can be controlled, for example, by varying the applied shear or normal force.

The particular adhesive material employed for these tests consists of arrays of 80μm long, angled silicone rubber "micro-wedges" fabricated on a 25μm thick polyimide film (Day *et al.*, 2013). The same adhesive film has been used in a passive gripper for lifting objects purely through shear tractions (Hawkes *et al.*, 2015) and in grippers designed to grasp objects in space (Jiang *et al.*, 2017). In the present application we use the film with a small positive normal pressure to

provide greatly enhanced friction. Because the gripping performance depends on both the area of contact and the pressure, we use a gripper equipped with a tactile sensing array. Many tactile sensing technologies are potentially applicable, with reviews provided in Cutkosky & Provancher (2016); Dahiya *et al.* (2010). Practical concerns include spatial and pressure sensing resolution, accuracy and robustness. For the experiments reported here we use a pre-commercial version of a tactile array reported in Le *et al.* (2017) with sensitivity down to normal pressures of 0.5 kPa. The sensor also includes a dynamic mode which, although not used in this work, provides a strong signal whenever contact first occurs. Though not designed specifically for this application, it performs well in the aforementioned metrics and was available for incorporation in the gripper without custom fabrication.

2.4 Theory

2.4.1 Friction with adhesion

As explained in Israelachvili (2015), friction generally has two components: one due to molecular attraction and hysteresis and one due to molecules bumping over each other. The former is an adhesion-controlled component, which depends on the real area of contact at a molecular scale. The latter is a load-controlled component, which depends on the normal force. For most hard materials, the former part is negligible and the latter part provides a maximum friction force that grows linearly with the applied load; hence $f_\mathrm{t} \leq \mu f_\mathrm{n}$, where f_t is the tangential force, f_n is the normal force and μ is the coefficient of friction. However, with gecko-inspired adhesives, even in the presence of a normal force, the area-dependent part often dominates.

Thus, in static conditions, we expect the tangential force for an adhesive with a positive normal force to follow:

$$f_\mathrm{t} \leq \int_A (\mathrm{c_l}\, p + \mathrm{c_a})\, \mathrm{d}A, \tag{2.1}$$

where $p(x, y)$ is the pressure at a given location of the contact, $\mathrm{c_l}$ and $\mathrm{c_a}$ are constants and A is the contact area. Similarly, in static conditions, the moment about an axis perpendicular to the

fingertip's surface should correspond to:

$$m_z \leq \int_A \|\boldsymbol{r}\|_2 \left(c_l\, p + c_a\right) \mathrm{d}A, \tag{2.2}$$

where $\boldsymbol{r} = [x, y]$ is a vector from the center of pressure of A to each element in A.

The work in this paper involves directional adhesives, hence the constant c_a becomes a function of the angle between the preferred loading direction of the adhesive and the angle of the applied tangential force: $c_a(\phi)$. For example, fig. 2.2 shows two different possible arrangements of directional adhesives.

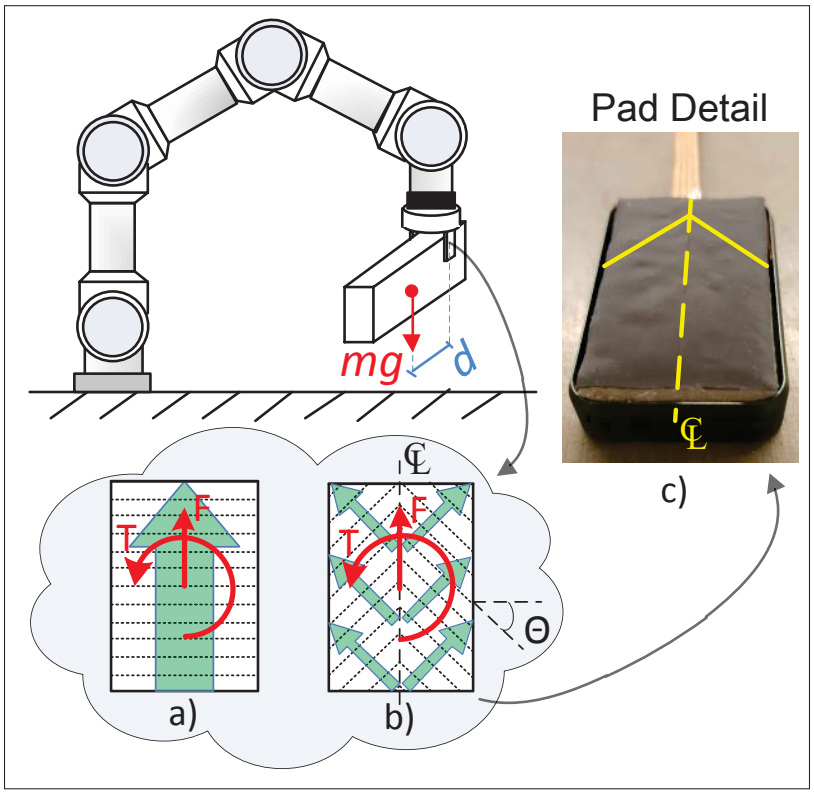


Figure 2.2 Drawing of a robot picking up an object away from its center of gravity (Bottom: Force and torque generated by a fingertip covered with a) uniformly aligned straight wedges ($\theta = 0°$) and b) a chevron pattern with wedges rotated $\pm\theta$ degrees c) Inset shows detail of gecko material on pad, arranged in chevron pattern matching b))

2.4.2 Moment compensation in manipulation tasks

If a gripper only needs to exert tangential forces in a single direction, as when lifting objects by grasping at their center of mass, it is most efficient to orient the directional adhesive parallel to the lifting direction. However, manipulation typically involves rotating grasped objects about multiple axes. In addition, the objects may be non-homogeneous, or it may not be possible to grasp them along their centerlines. Consequently, grasp attempts with a single manipulator often introduce a moment about the point of contact with the object, as shown in fig. 2.2. With a tight pinch of the object this is perhaps not a problem. However, when attempting to minimize grasping forces there is the potential for slippage. A moment introduces a circular pattern of shear forces around the center of rotation, of which only a small portion is aligned with the adhesive's strongest direction.

To compensate for the aforementioned effect, one can arrange small areas of directional adhesive with multiple orientations. Fig. 2.3 shows the expected results for several different patterns, assuming an adhesive that has maximum strength at $\phi = 0$ and minimum strength in the orthogonal directions at $\phi = \pm 90^\circ$. The color scale of the plots is distributed linearly with ϕ, which is defined as the angle between the adhesive's preferred direction and the actual loading direction (see fig. 2.3d).

It is also noteworthy from fig. 2.3 that θ should in general be adjusted according to the fingertip's dimensions. Furthermore, for optimal moment compensation on a rectangular surface, regardless of tangential force compensation, one should orient the wedges such that:

$$\theta_m = \tan^{-1}\left(\frac{\mathrm{h_f}}{\mathrm{w_f}}\right), \tag{2.3}$$

where $\mathrm{h_f}$ and $\mathrm{w_f}$ are respectively the height and the width of the fingertip's contact area. Choosing θ accordingly ensures the wedges are parallel with the fingertip's principal diagonals.

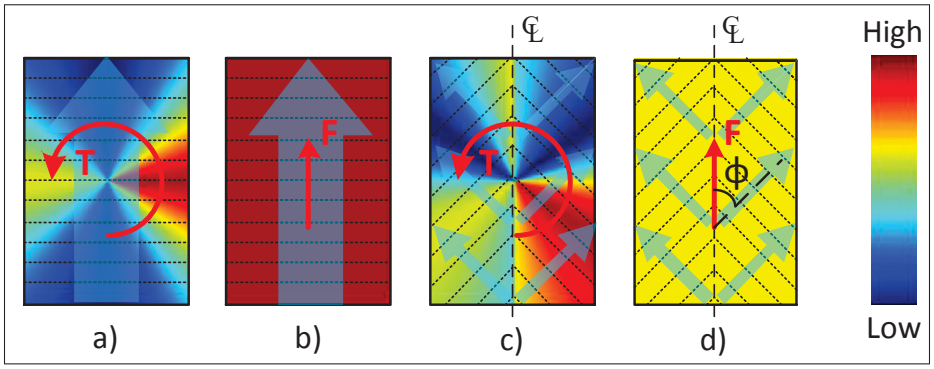


Figure 2.3 Efficiency map comparison of a straight (a-b) and an oriented (c-d) gecko adhesive designs when generating a pure counterclockwise torque (a-c) or pure upward tangential force (b-d) (Semi-transparent arrows indicate the preferred loading direction of the gecko adhesive)

2.5 Fingertip design and construction

The gripper consists of an industrial Robotiq® Two-Finger 85 gripper equipped with tactile sensors. The sensor is described in Le *et al.* (2017). This multi-modal tactile sensor is 22mm x 42mm and contains a 4-by-7 matrix of taxels. The outer skin of the sensor was replaced by a film with patches of gecko-inspired adhesives in one or more directions (see fig. 2.2 for details). As noted earlier, the adhesive used in these experiments is the same as described in Day *et al.* (2013); Jiang *et al.* (2017) and consists of rows of inclined "micro-wedges" with a triangular cross section. A characteristic of this adhesive is that the amount of adhesion increases in proportion to the applied shear load. When the shear load is relaxed, the adhesion becomes negligible.

2.6 Experiments

Before conducting robotic grasping experiments we conducted preliminary bench-top experiments to characterize the adhesive material and compare it with conventional silicone rubber gripper pads.

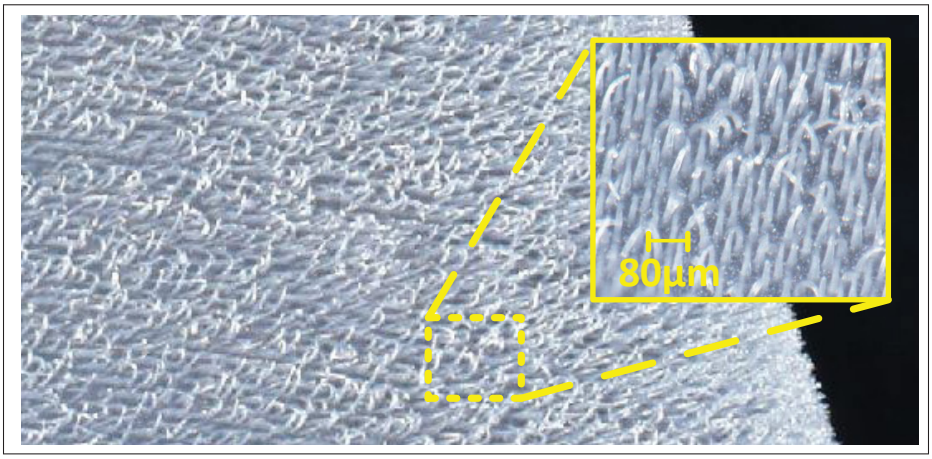


Figure 2.4 A microscopic view of the synthetic hair manufactured for comparison purposes (average hair length is approximately 80 µm)

2.6.1 Friction comparison with flat and textured silicone rubber

2.6.1.1 Experiment

Pull tests were conducted for the directional adhesive, for a non-directional and non-adhesive texture of 80 µm long silicone hairs (Fig. 2.4), and flat silicone rubber. In each case the material was Sylgard® 170 PDMS.

In these tests we affixed samples of the material to a manually adjustable stage with a power screw turned by a crank. We used calibrated weights to adjust the normal force and measured the tangential forces using a Mark-10 M4-10 force gauge (0.02N resolution) while turning the crank.

2.6.1.2 Results

Fig. 2.5 compares the three silicone rubber materials. The gecko adhesive shows dramatically higher tangential force at very light normal force. Beyond 20 N all three materials show an approximately linear increase with normal force. The magnified portion of the graph shows that at light loads, the nondirectional array of 80 µm hairs achieves a more uniform contact and produces a higher tangential force, despite having less theoretical contact area due to gaps

between the hairs. This result is consistent with observations that compliant and textured rubber surfaces work better than solid rubber at light loads (Cutkosky *et al.*, 1987).

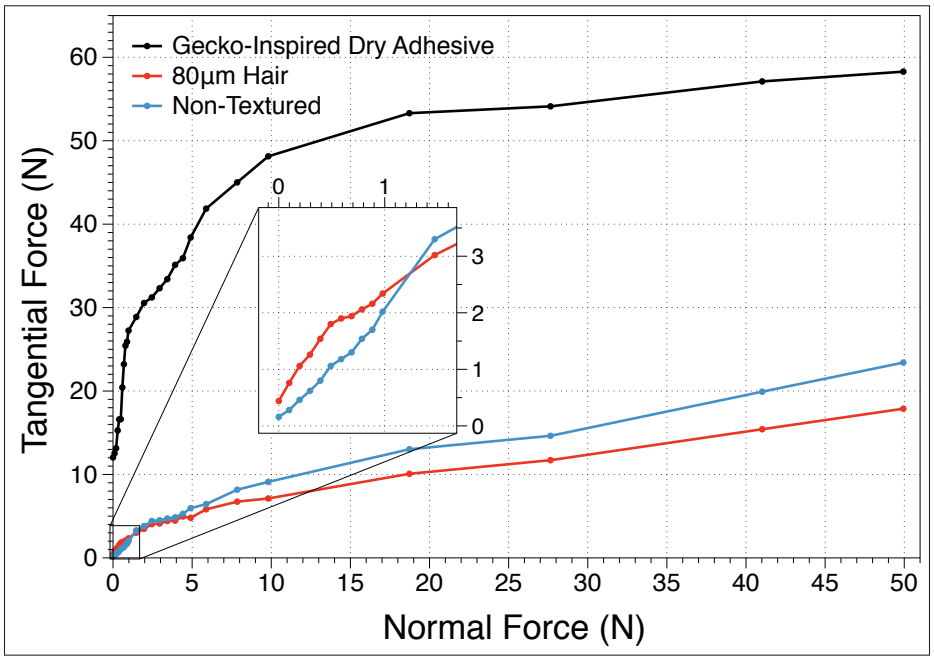


Figure 2.5 Comparison of maximum tangential force, $F_{t,max}$ as a function of normal force for 31 mm^2 samples of (i) gecko-inspired directional adhesives pulling in the preferred direction, (ii) a texture of non-directional silicone hairs (fig. 2.4), and (iii) flat silicone PDMS

2.6.2 Tangential force and area versus normal force

2.6.2.1 Experiment

According to eq. (2.1), the maximum tangential force for a sample of gecko-inspired adhesive should depend on both the real area of contact and the normal force. However, there are a couple of factors that make prediction difficult. First, the real area of contact, over which the adhesive and the adherend surface are within molecular distances, is typically much less than the nominal area of contact ($A_r << A$). Second, the real area of contact will typically increase with increasing normal force, i.e., $A_r \rightarrow A$ as f_n becomes very large.

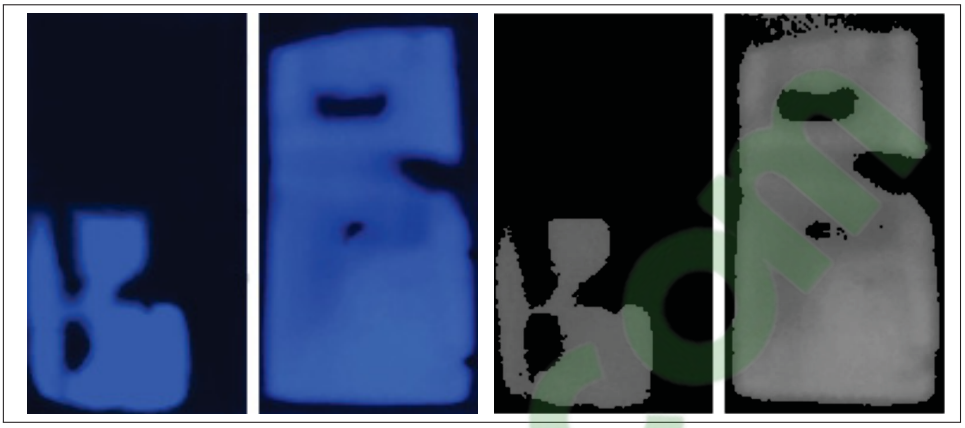


Figure 2.6 Example FTIR images for low and high pressure conditions: Left two images are shown before image processing and right two images are post processing – bright regions correspond to adhesive contact

To explore these effects in more detail we measured the growth of the area of contact and the maximum tangential force for increasing amounts of normal force. Following the approach described in Eason *et al.* (2015), frustrated internal reflection (FTIR) imaging provides an estimate of the area of contact. Fig. 2.6 shows typical FTIR images for low and high areas of contact, respectively. To obtain results consistent with those expected in robotic grasping, we conducted these experiments using an adhesive sample on a glass plate, pressed against one jaw of the robot gripper shown in fig. 2.1. With the opposite jaw of the gripper removed, there was room for a camera to record the FTIR image. Using Matlab's native image processing functions, we rendered the images of the contact to a grey-scale intensity map and then integrated the pixel by pixel intensity of the image across the gripping surface. This method yielded the percentage of maximum light intensity of each trial, which was then plotted against normal force applied. Note that while this method does not necessarily give an accurate absolute estimate of A_r, it does allow us to track the growth of A_r with increasing normal force.

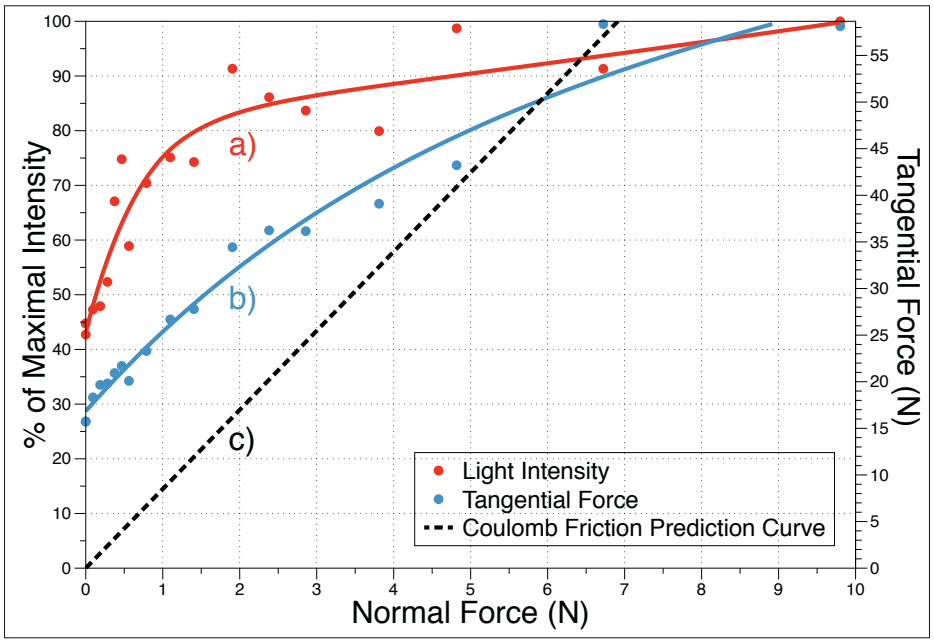


Figure 2.7 (a) Normalized contact area measured using FTIR imaging and (b) maximum tangential force ($\phi = 0$) as a function of increasing normal force, F_n. (c) Coulomb friction is not a good match, especially at low pressures. However, F_{tan} is also not simply a linear function of area

2.6.2.2 Results

Fig. 2.7, as anticipated, shows that the maximum tangential force and area of contact increase nonlinearly with increasing normal force. The plotted line (c) shows that a Coulomb friction model fits the data poorly, especially at light normal forces, and confirms that the adhesion-controlled component of the shear force is significant. In addition, it is clear that the maximum tangential force is not exclusively a function of the area; therefore the load-controlled component of the shear force cannot be ignored either.

2.6.3 Effect of tangential force direction

2.6.3.1 Experiments

We repeated the tests from Section 2.6.1 while pulling the gecko-inspired adhesive at various angles, ϕ, with respect to its preferred loading direction.

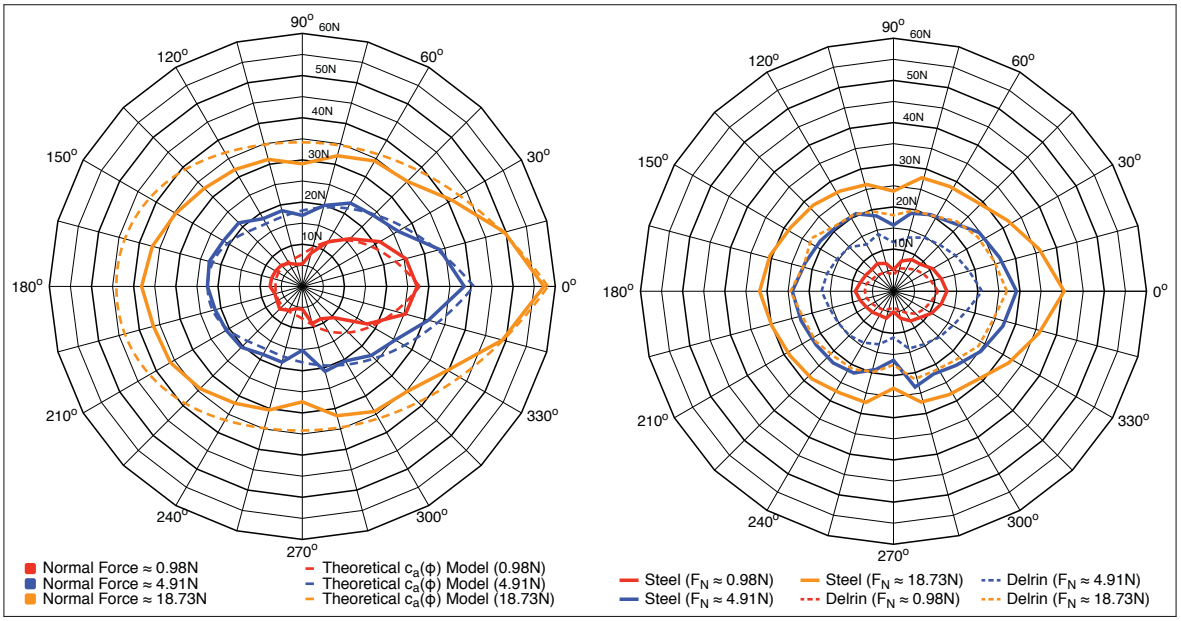


Figure 2.8 Maximum tangential force as a function of the pulling angle (ϕ), for three different normal forces: upper plot is for smooth acrylic, lower plots are for steel bar stock and machined Delrin®

2.6.3.2 Results

Fig. 2.8 illustrates the dependence of the maximum tangential force on pulling direction for three values of normal force, ranging from a gentle gripping force of 0.98 N to a firm grip of 18.7 N. Tests were conducted on smooth acrylic sheet, cold rolled steel bar stock ($\approx 1.17\mu$m Ra roughness) and machined Delrin® acetal homopolymer ($\approx 0.55\mu$m Ra roughness). In each case the maximum tangential force is approximately twice as high in the preferred direction as compared to the least effective direction. Note that pulling "backward" ($\phi = 180°$) is not

the worst case because the microwedges can flip over and provide some adhesion, albeit with a smaller contact area than in the preferred direction. Hence the least effective loading direction is $\phi = \pm 90°$. Also shown is an approximating model for $c_a(\phi)$, using a third-order polynomial, as discussed in section 2.7.

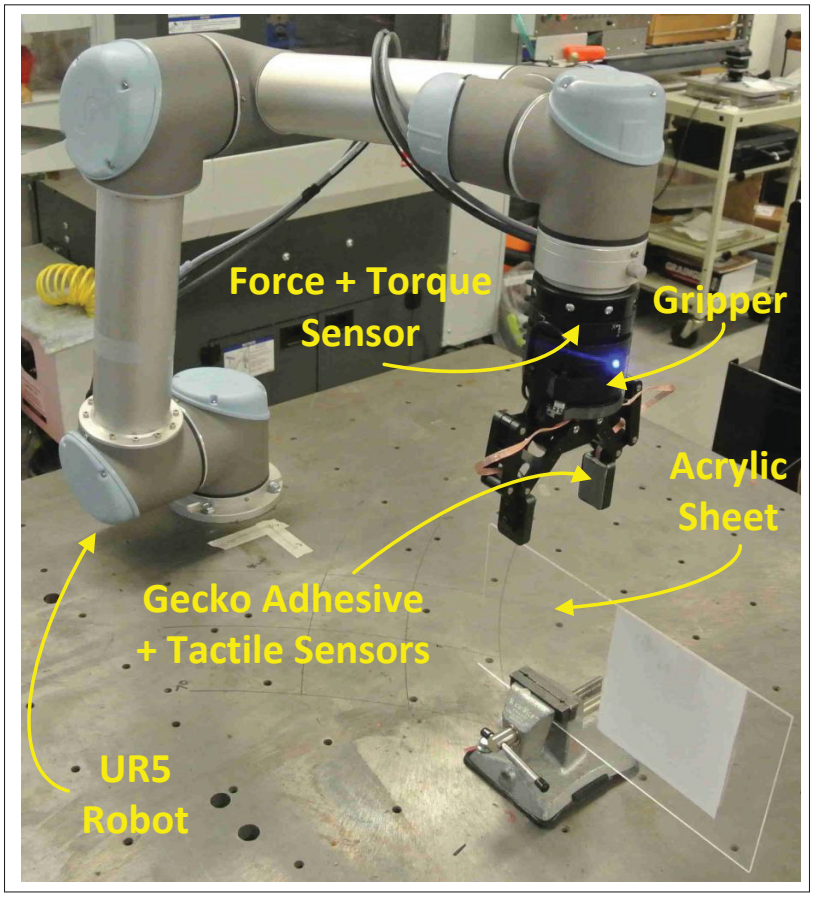


Figure 2.9 Robotic grasping experiment setup

2.6.4 Robotic grasping experiment

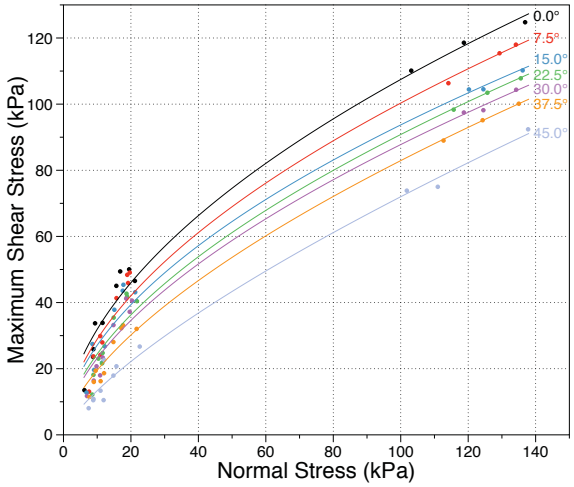
2.6.4.1 Setup

The experimental setup shown in fig. 2.9 was used for this experiment. A UR-5 robot arm was equipped with a Robotiq Two-finger 85 gripper and an FT-300 force/torque sensing wrist. As noted earlier, each gripper finger had a 7-by-4 tactile sensor array (Le *et al.*, 2017) and a

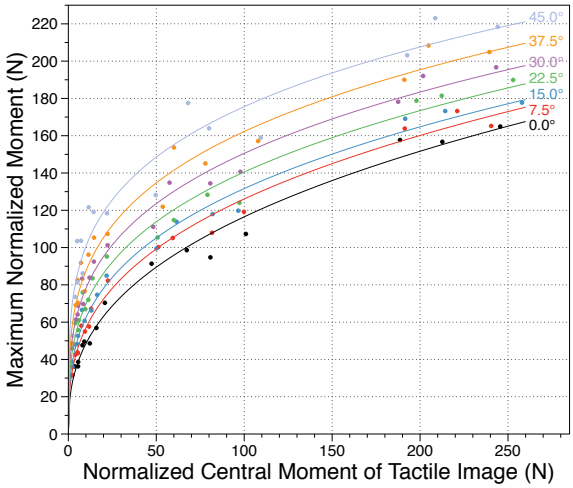
patterned skin of directional adhesive. The data from these various sensors were combined with position and velocity information from encoders in the robot arm to provide a complete picture of the dynamic and static state of the robotic grasp.

For this round of testing, a set of specialized adhesive surfaces were made to achieve the diagonal "chevron" pattern of micro-wedges introduced in section 2.4. This pattern strikes a useful balance between resistance to moments and ease of fabrication.

Multiple panels of gecko adhesive were made with the orientation angle of the two halves varying in 7.5° increments from $\theta = 0°$ to $\theta = 45°$, i.e. from completely perpendicular to line of force to 45° away from the line of force. Next, various experiments were run using each set of gecko-adhesive covered panels. Two motion patterns of the robot arm were devised. Each was run with progressively increasing nominal surface contact on the acrylic plate, starting at 25%, then going to 50%, 75%, and finally 100% coverage (regulated by hand measurement and preprogrammed grasp points). All started by pinching a rigidly-attached sheet of acrylic approximately one quarter inch thick with a portion of the adhesive-coated two-finger gripper at a specified level of normal force. During the tests, this pinching force was also varied by modulating the gripper's closure set-point. Unfortunately, this discretization of the available set point values was such that we could only vary between four different normal forces before saturating actuators and sensors, and only three of those can reasonably be termed as falling into the low pressure range. After this pinch had been achieved, the first test started by pulling on the rigidly attached acrylic sheet in an upward direction, such that a pure shear force was exerted on the fingertips' adhesive surface. Depending on the normal force and the contact area, slippage occurred at a certain level of tangential force, which was recorded. The second motion pattern consisted of a rotation around the contact surface's centroid, such that a pure moment was exerted on the fingertip's adhesives.



a) Motion 1) Maximal shear stress (τ) as a function of normal stress (σ) for different values of θ: here, for each angle's dataset, a power function ($y = ax^b$) was fit to the experimental data to show general trends



b) Motion 2) Maximal normalized torque (M') as a function of normalized normal stress (μ) for different values of θ: a series of power functions were again fit to the experimental data to show the general trends

Figure 2.10 The robotic grasping experiment results

2.6.4.2 Results

With reference to equations 2.1 and 2.2, we can provide a discretized result for the normal and shear stress, corresponding to the number of elements, N, in the tactile array, each of area a_t:

$$\sigma = \frac{\sum_{i=1}^{N} P_i}{N\, a_t} \tag{2.4}$$

$$\tau = \frac{F_T}{N\, a_t} \tag{2.5}$$

where P_i is the force count at each taxel and F_T is the tangential force, as measured by the force-torque sensor located in the wrist. Fig. 2.10a shows the relation between maximum shear stress and normal stress for each set of patterned adhesives that was tested. As expected, the greatest tangential force support was generally obtained with the non-angled wedges (in this case, $\theta = \phi = 0°$). The fit power functions clearly demonstrate the important trends from the dataset: as θ grows, the maximum allowable shear stress decreases consistently, showing the effect of varying the wedges' orientation.

Additionally, the normalized central moment of a tactile image is given by:

$$\mu = \frac{\sum_{i=1}^{N} \sqrt{x_i^2 + y_i^2} P_i}{\frac{1}{N} \sum_{i=1}^{N} \sqrt{x_i^2 + y_i^2}} \tag{2.6}$$

where $\{x_i, y_i\}$ are the coordinates of taxel i with reference to the centroid. The numerator of (2.6) is the sum of each taxel's contribution to the total measured moment at the wrist, while the denominator is the averaged distance between the taxels and the centroid and normalizes the numerator to allow better comparisons among the data having different contact area values. Similarly, the normalized moment is given by:

$$M' = \frac{M}{\frac{1}{N} \sum_{i=1}^{N} \sqrt{x_i^2 + y_i^2}} \tag{2.7}$$

where M is the moment measured by the force-torque sensor. Fig. 2.10b shows the relation between the maximum normalized moment and the tactile map's normalized central moment for different values of θ. Contrary to the previous case (fig. 2.10a), the maximum allowable moment grows with θ, which is a result that was anticipated from section 2.4.2. The non-linear nature of the curves is also expected, given the dependence that the real area of contact (A_r) has on pressure, as discussed in section 2.6.1.

2.7 Discussion: predicting maximum shear stress in practical settings

In this section, we discuss a practical method to derive an empirical model that predicts the performance of a gripper such as the one used here, with a minimal set of data acquired during a calibration phase. The real area of contact being a function of both pressure and loading angle (figs. 2.7 and 2.8) explains the non-linear nature observed in the trends and the differences seen in curve shape across the different trials (fig. 2.10). The real area of contact changing with pressure is probably due to surface deformation causing some areas of the surface to make contact before other portions as seen in fig. 2.6. A quantitative model of that deformation would require a detailed elastic model of the gripper pads, which is beyond the scope of this paper.

Moreover, such an analysis would be rendered mostly obsolete as soon as the gripper-adhesive connection is altered in any way.

Furthermore, although it is believed that the impact of pressure and loading angle are physically related on the micro-scale, our data indicated that the effect each has on shear strength can be assumed independent at low pressure values without significant loss of accuracy. Thus, a simple model can be generated by decoupling pressure and loading angle effects into two separate terms: a positive component for shear stress as a function of pressure and a negative term that is subtracted depending on how far the loading is from the ideal loading angle. Then, relying on a calibration data set composed of a pressure data set from section 2.6.4 (i.e. data with $\phi = 0°$, shown in green in fig. 2.11) and a set of data from the pulling angle experiments (section 2.6.3), we can use curve-fitting tools to fit a proper function to the terms:

$$\tau(P, \phi) \approx f_1(P) + f_2(\phi). \tag{2.8}$$

Through the use of this calibration phase, a mathematical relation is thus generated and calculates maximal shear strength as a function of both pressure and loading angle. An example of such a function $\tau(P, \phi)$ is plotted as a surface in fig. 2.11 with the actual data points from our experimental data overlaid in their appropriate locations. In this example, $f_1(P)$ was empirically fitted to a double exponential function, while $f_2(\phi)$ is a third-order polynomial. The mean error between the theoretical model and the real measured values is $\mu \approx 2.97$kPa, with a standard deviation of $\sigma \approx 2.67$kPa.

Given the generally good agreement between data points and the predictive surface, we believe this estimation has potential to be useful as a predictive tool. At higher pressure values, however, the accuracy of the model starts to deteriorate. Indeed, for pressures below 80 kPa the error's mean and standard deviation are respectively $\mu_L \approx 1.46$kPa and $\sigma_L \approx 1.26$kPa, while for pressures above 80kPa, $\mu_H \approx 8.13$kPa and $\sigma_H \approx 7.39$kPa. A likely explanation for this effect is the decoupling of pressure and loading angle effects, which is justified only at low pressures.

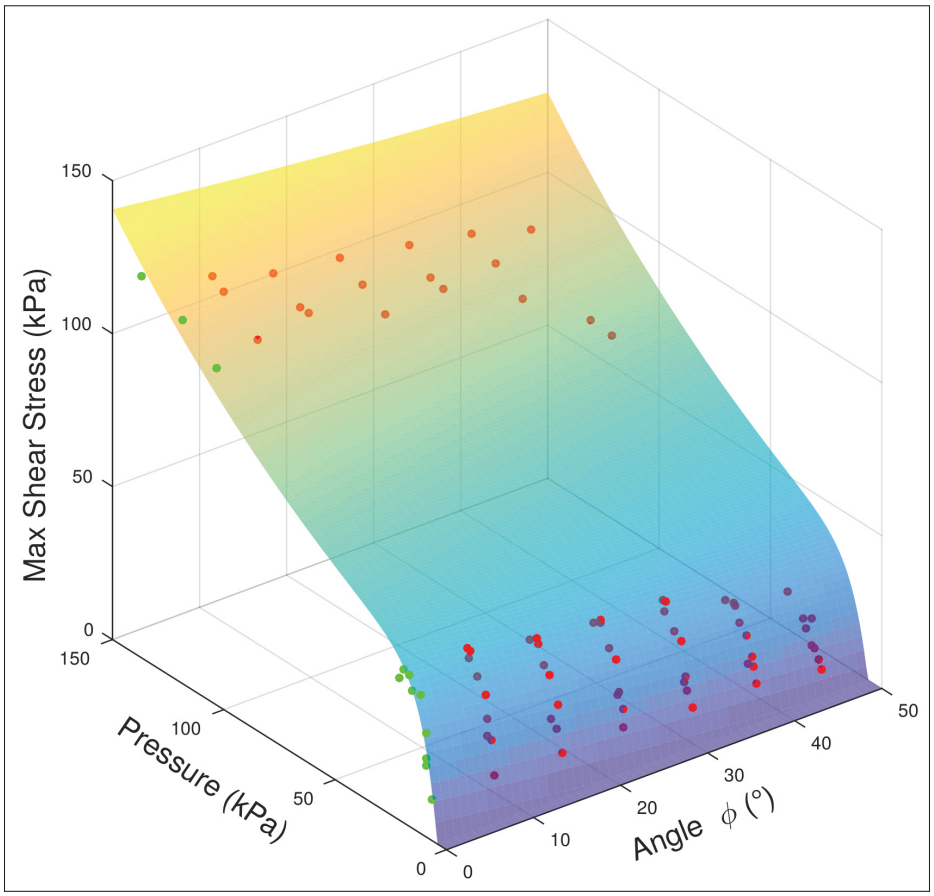


Figure 2.11 Expected shear stress vs normal stress and loading angle: the green points represent original data points used for calibration and the red points represent real data points that the surface is intended to predict

At higher normal forces, we believe this decoupling is less representative of the actual system because high pressures crush the wedges into contact with the surface no matter the loading angle, producing a consistently high real area of contact. Fortunately, this deterioration of the model's accuracy is mostly limited to those high pressure conditions, which are of less practical interest in this gentle manipulation focused application and for which a Coulomb friction model is increasingly valid on its own.

2.8 Conclusions and future work

2.8.1 Conclusions

As originally hypothesized, the experiments have confirmed that there are three primary parameters to vary when using a gecko-inspired adhesive equipped gripper. Normal force has a strong positive correlation with maximum shear force that can be exerted, but it is also the parameter that needs to be minimized. Thus, one can turn to loading angle, ϕ, and real area of contact to achieve the desired shear strength.

The experiments have shown that real area of contact is not necessarily the same as the macro-scale contact. The non-linear trends seen in our experiments consistently have a steeper increase in shear strength and area for the initial, low normal force values. Hence, it is recommended that small, but intentional pressure be maintained with these and similar grippers to maximize usage of the gecko-inspired adhesive.

Additionally, we show that even small deviations in loading angle have a noticeable impact on maximum shear force and that the chevron pattern proposed has the anticipated benefits for resisting torsional loads with commensurate loss of performance when resisting pure shear conditions. This trade off will generally have to be decided on an application by application basis. In general, however, single manipulators may see better overall efficiency with oriented adhesive patches (with $\theta \neq 0$) as their single point of contact will almost always introduce a non-trivial moment in any given task. For situations where cooperative manipulation is possible, the straight wedges will likely emerge as the better overall choice because necessary torques can be generated through the right combination of point forces and do not need to be created using the end effector's contact area itself.

Further, accurate prediction of a given gripper's performance characteristics is possible with a relatively small set of calibration data. At least one data set needs to be taken to determine the relationship between pressure and shear strength. This may be sufficient when combined with the loading angle relationship shown in this paper, but if the grasped material has significantly

different surface properties than the acrylic, steel and Delrin tested, a single round of testing similar to section 2.6.3 could be performed to improve accuracy and fully inform the predictive model proposed.

Lastly, with the knowledge of anticipated gripper performance, accurate application of grip strength can be combined with the improved contact properties of the gecko-inspired adhesive itself to significantly reduce the overall demand on gripper actuators. Smaller actuators can lead to lighter, safer tools on robotic arms, which, in turn, could be highly beneficial to collaborative robots.

2.8.2 Future work

There are three immediate directions in which this work should be extended. First, additional testing is needed to confirm that the observed trends extend to additional materials and roughnesses. Second, working outside the ideal condition of a flat grasping surface will be necessary to see how significantly performance drops in such scenarios. Third, dynamic tests will eventually be necessary to show that the results produced here can be applied to complex manipulation tasks beyond the quasi-static conditions on which these tests were focused.

2.9 Acknowledgments

We thank Elliot Hawkes and Arul Srinivasan for their suggestions regarding the combined friction/adhesion model in this paper. This work was funded in part by a grant from Ford Motor Company. Jean-Philippe Roberge is supported by an NSERC scholarship. Wilson Ruotolo is supported by an NSF GraduateResearch Fellowship.

CHAPTER 3

IDENTIFYING IMPORTANT ROBOTIC EVENTS USING SPARSE TACTILE DATA REPRESENTATIONS

Jean-Philippe Roberge[1], Tony Wong[1], Vincent Duchaine[1]

[1] Systems Engineering Department, École de Technologie Supérieure,
1100 Notre-Dame Ouest, Montréal, Québec, Canada H3C 1K3

Paper submitted to the IEEE Transactions on Robotics, July 2019.

3.1 Abstract

During manipulation tasks, it is often necessary to detect the occurrence of important dynamic events that could affect, or sometimes even jeopardize the ongoing robotic operation. Besides, having the ability to properly identify the nature of the said events is also crucial to many different contexts. For instance, a firmly-grasped object that makes contact or slides across an external surface might not always be problematic, while a loose object starting to slip out of the gripper is generally more concerning. In this paper, we address an advanced tactile perception problem that consists of detecting, as well as distinguishing between four different types of such important events. To do so, we rely solely on a spectral analysis of the vibrations emitted during the events, which are captured using tactile sensors. We propose to use sparse coding as an unsupervised learning technique to build a dictionary of high-level tactile features that can be used to detect and distinguish the classes of events. During online operations, only a small subset of the features that best represent the vibration measurements is sufficient to accurately determine what event is taking place. Furthermore, we show that once the features are learned using easily-generated unlabeled data, they subsequently generalize very efficiently to other sensor technologies. We present an in-depth analysis of each of the main sparse coding hyperparameters' effects on the system's efficiency, and draw conclusions about their influence on the signals' representation quality. Our method was tested on data obtained from a total of more than 1,600 experiments that were conducted on 62 different everyday objects. Results

show that using our approach, it was possible to almost perfectly distinguish and detect all of the considered dynamic events with an almost perfect success rate of **99.90**%.

Keywords: Tactile sensing, tactile events, grasping, sparse coding.

3.2 Introduction

Rather than performing a small set of tasks in a repetitive manner as they used to do in the past, today's robots are asked to execute increasingly-complex and varied operations. In addition, the growing adoption of collaborative robots worldwide brings new challenging requirements as these robots are more and more operating in the same workspace as humans.

In order for robots to adapt to the new role they are asked to fulfill, they will need to significantly improve their ability to perceive and interact with their immediate environment. Artificial vision has been considerably developed during the last decades, giving robots the ability to represent their environment accurately, which in turn allowed them to be more efficient at a large set of industrial tasks. However, the vast majority of today's industrial robots are still not equipped with sensors that provide them an artificial sense of touch, although this sense is possibly at the heart of their interaction capabilities. One of the reasons for these circumstances is the rather large gap to bridge between the complex raw data generated by tactile sensors and the proper way to translate these data into enhanced robotic manipulation skills.

Nevertheless, several examples of recent work (Yussof *et al.*, 2009; Romano *et al.*, 2011; Dang & Allen, 2014; Shirafuji & Hosoda, 2011; Roberge *et al.*, 2016) have highlighted how beneficial tactile sensors are to different tasks when such sensors are integrated to robotic grippers. An extensive review of tactile sensing transduction techniques and tactile sensing in robotic applications is made in Kappassov *et al.* (2015). As noted in Dahiya *et al.* (2010), a bigger proportion of tactile-sensing-related papers deals with *static*-based tactile sensing, which is a modality that uses *static* pressure sensing data, rather than *dynamic*-based sensing, which concerns fast-evolving dynamic signals such as vibration measurements.

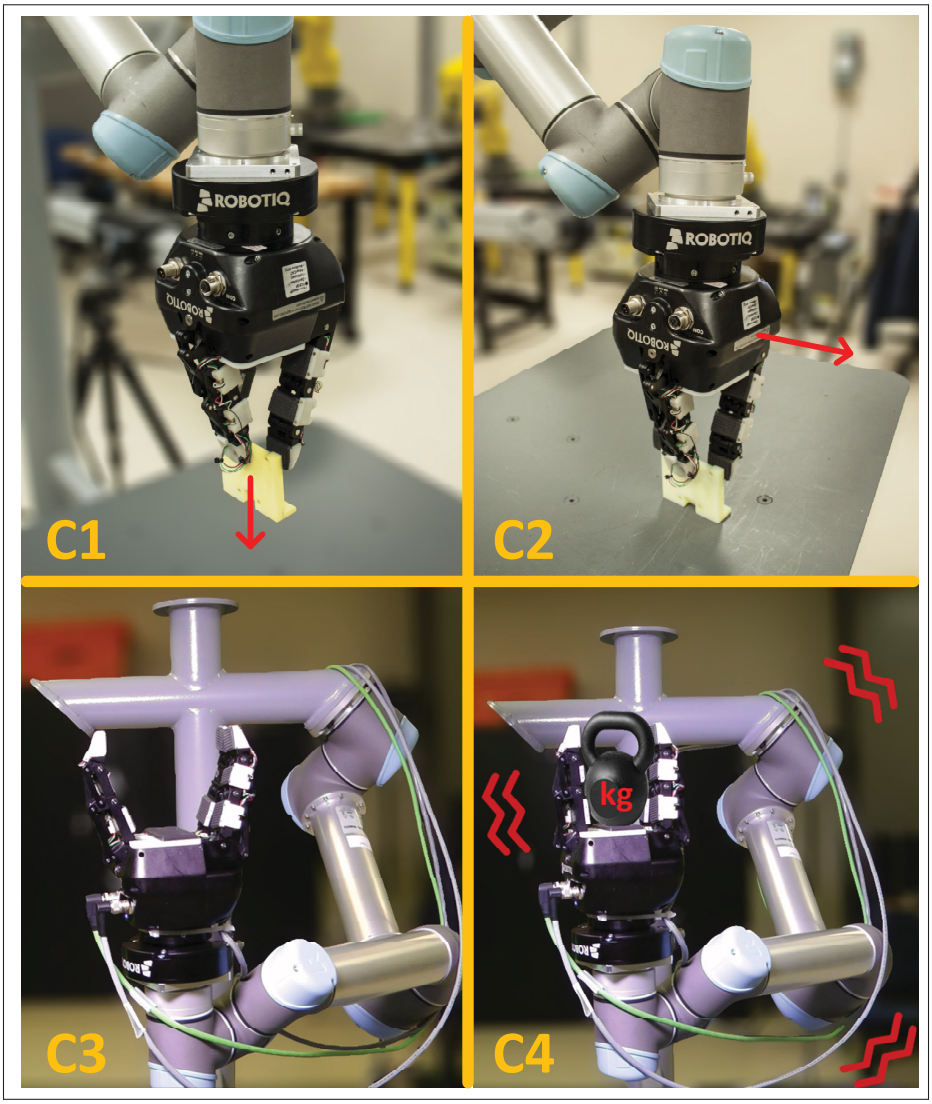


Figure 3.1 The four classes of tactile events considered during this work: C1 - An object slipping inside the gripper; C2 - An object slipping across a surface; C3 - Null Class (no relevant vibrations are occurring) and; C4 - External Vibrations (here, generated by the UR5 robot while lifting a load surpassing its payload limit)

The *static* modality is well-fitted for quantifying the force exerted on an object (Dang & Allen, 2014), or for getting information about the configuration of an object held in the gripper. For example, Madry *et al.* (2014) use static pressure information to recognize objects, while also assessing the stability of the grasp. Cockburn *et al.* (2017) rely solely on static pressure encodings to assess grasp stability, while Kwiatkowski *et al.* (2017) also use static pressure data, but now

fused with proprioception to perform the same task. While these examples show how static tactile sensing can provide useful information for grasping and manipulation, they still only concern slow-evolving contexts and do not address how to provide information to a robot about the rapidly-evolving events that happen during operation. Yet, this information is essential to develop efficient reflexes that would allow robots to better interact with their environment in the future.

Indeed, when a robot is interacting with its environment, one of the main concerns is rather its ability to quickly assimilate important *dynamic* events as they occur, events for which *dynamic* data are better suited. For example, given the critical nature of in-hand object slippage, its reliable detection has been a concern for roboticists for a long time. Attempts to properly detect slippage are discussed in Melchiorri (2000); Holweg *et al.* (1996); Tremblay & Cutkosky (1993); Wyk & Falco (2018); Zhang *et al.* (2018), with reported success rates ranging from around 75% to 97%. Among slippage-detection-related literature, some have also successfully distinguished linear slippage from rotational slippage (Melchiorri, 2000; Su *et al.*, 2015; Meier *et al.*, 2016b).

On the other hand, only few works have tackled the problem of detecting and classifying different dynamic events other than the ones related to in-hand object slippage. The latter is a difficult classification problem to solve, furthermore, tactile signals that are visually similar might still be related to fundamentally different dynamic events, making it complex to handcraft features to discriminate the said classes. Goeger *et al.* (2009) classify tactile data into three classes: in-hand slippage, no slippage and other dynamic events. While this work's content is innovative and interesting, the authors got a very good classification rate only for one out of three classes. Meier *et al.* (2016a) proposes to use a shallow convolutional neural network (CNN) to distinguish in-hand object slippage from object on-surface sliding and get a classification success rate ranging from 76.25% to 81%. Those networks can learn to properly encode the data stream while also learning to classify them. However, even a shallow CNN has a lot of hyperparameters to learn and thus requires a large quantity of labeled data. For example, in the latter work, the network had several hundreds of hyperparameters to learn, but only 125 training data were acquired using a small variety of objects (only three were available), which might limit the

efficiency and scalability of such an approach. Heyneman & Cutkosky (2016) use handcrafted features from the tactile signal's power spectral density (PSD) function to distinguish in-hand and object-world slippage. None of the aforementioned works have studied how well their method could generalize to other sensor and gripper technologies.

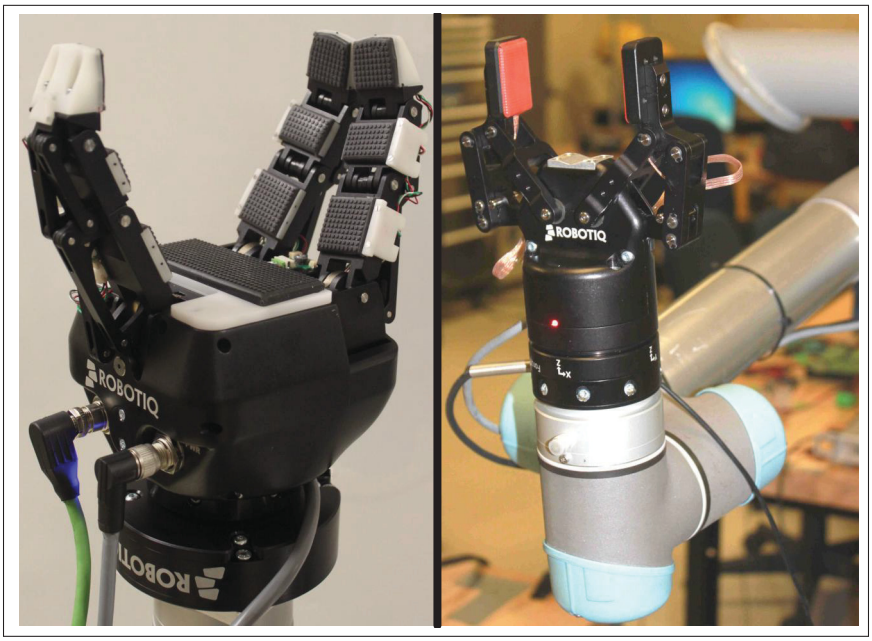


Figure 3.2 The two setups used during the experiments: a) on the left, a Robotiq 3-finger gripper equipped with our lab's *1st-gen* tactile sensors mounted on a UR5 robot and b) on the right, a Robotiq 2-finger gripper equipped with our *2nd-gen* tactile sensors, mounted on a different UR5 robot

In contrast, in a previous work (Roberge *et al.*, 2016) we introduced a promising method to disentangle the important high-level features from a raw dynamic signal, which was based on sparse coding (Olshausen & Field, 1996). While it was shown that using sparse tactile data encodings is useful for distinguishing dynamic events, yet no in-depth look was taken on the system to study how to tune the important parameters that significantly alter its performance, which is unsatisfactory to future implementers. Indeed, no investigations about the hyperparameters' impacts on the representations' quality and classification rates were carried out. On the contrary, the parameters were almost all set by trial and error. Also, the results were obtained only from

preliminary experiments that involved a relatively small amount of data (244) from the same amount of experiments made with 32 everyday objects. Most importantly, no demonstrations about if and how well the proposed approach could generalize to other sensors, robots and objects were done, which strongly limits the potential adoption of such an approach.

In this present work, these shortcomings are corrected and a new system that automatically and efficiently matches a dynamic signal to its corresponding class of dynamic event is introduced. Here, we aim at performing advanced tactile perception and consider no less than four classes (which are depicted in Fig. 3.1): 1) *Object-gripper slippage*, where an object is slipping inside the gripper, 2) *Object-world slippage*, where an object firmly-held by the gripper is slipping across a surface, 3) *Null class*, where no perceptible vibrations are currently happening and 4) *External Vibrations*, where the vibrations that are perceived by the tactile sensor are not related to slippage but to another event. To test our approach, we conducted $1,607$ experiments involving more than 62 different everyday-objects (some of which are shown in Roberge *et al.* (2016)). Section 3.3 provides theoretical details about the proposed approach and section 3.4 describes the robotic hardware and software components that were used to collect the data. In section 3.5, we present an analysis of the hyperparameters' impact on the data representations' quality and demonstrate how efficiently this method can generalize to other sensors and grippers. We validate our method by performing 10-fold cross validations on the data collected during our experiments, and show the classification accuracy for different scenarios. Finally, in section 3.6, we discuss the results and the future work they might inspire.

3.3 Proposed approach

As we seek to achieve our goal of distinguishing between the four event classes, one of the most important considerations is finding a data representation that will illuminate the high-level features that are the most relevant for classifying the events. With this idea in mind, the current section explains how we transformed a dynamic signal of a single distal tactile sensor for use as an input to our algorithm. The dynamic signal in question is a fast time-varying tactile signal sampled at $1,000$Hz, which is a direct function of the fingertip's instantaneous variation of

normal stress. A more detailed presentation of the tactile sensor will be given in the next section. Also, a general overview of the proposed approach is shown in Fig. 3.3.

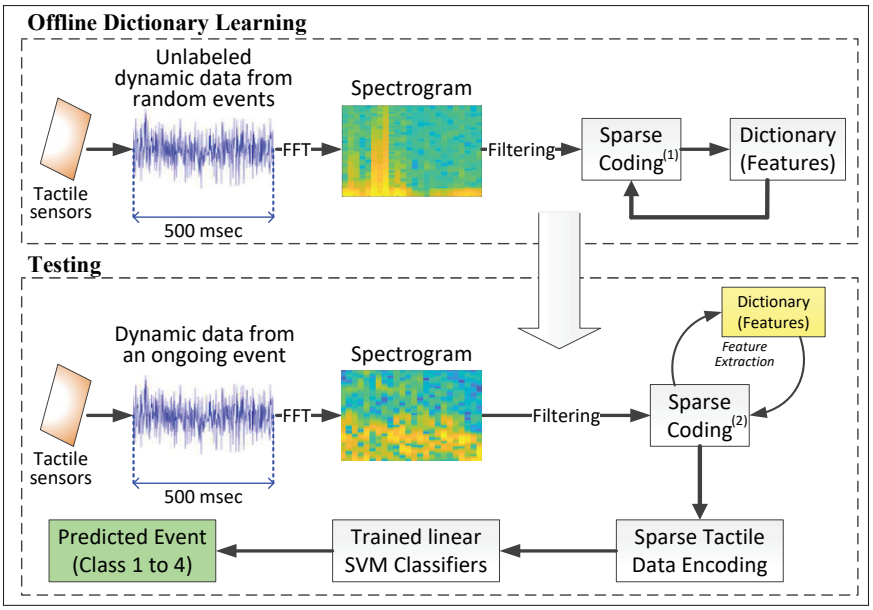
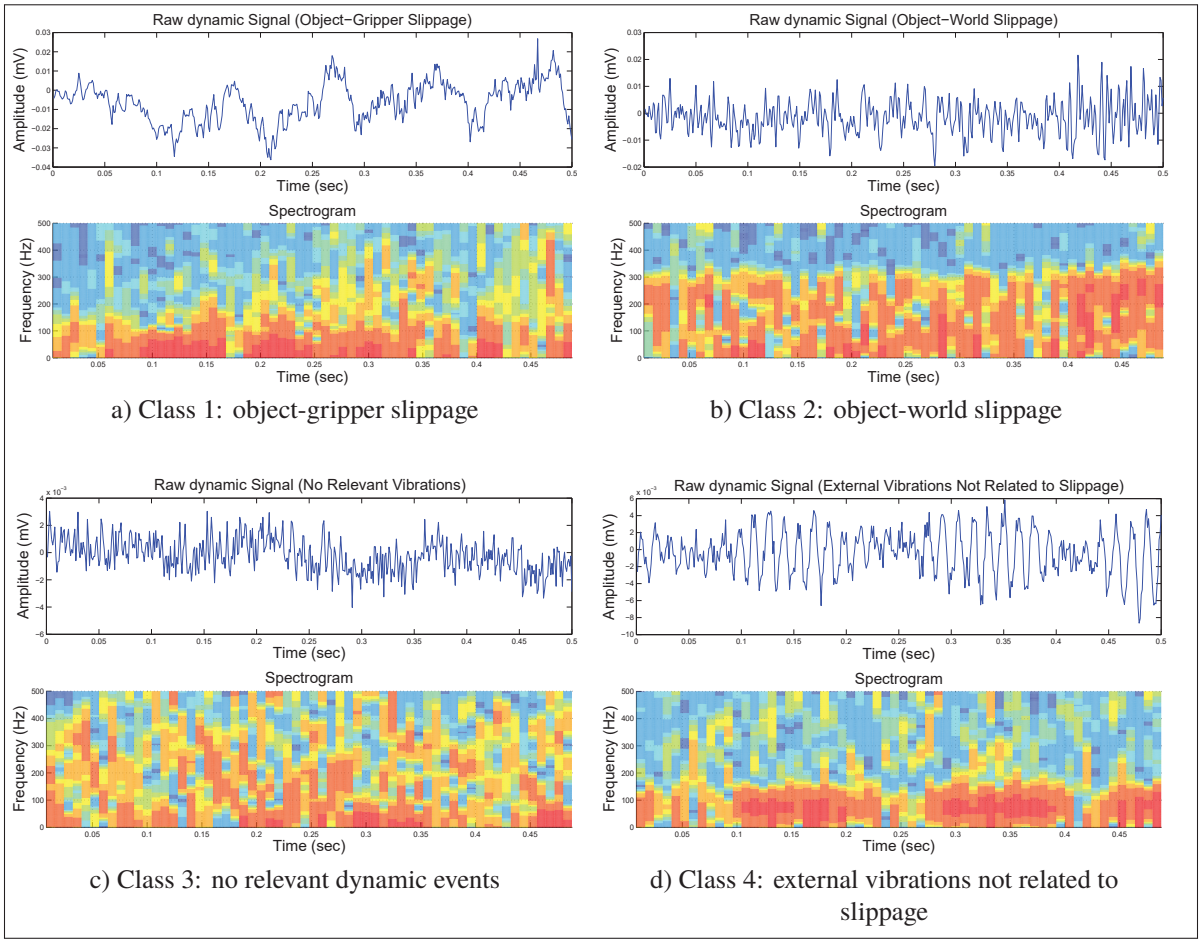


Figure 3.3 Functional diagram of our proposed approach: 1) The first sparse coding algorithm involves a dual optimization problem, see eq. 3.2 2) Once the dictionary is learned, the sparse coding algorithm is reduced to a single-optimization problem, see eq. 3.3

We will begin by further explaining the events in class 4, which encompass perceptible vibrations that are, on the other hand, not related to slippage. One example of the latter class, is when the robot (UR5) is undesirably vibrating while it is under certain specific conditions. Indeed, while experimenting with this apparatus, we found that using objects heavier than 2.5 kg sometimes caused the robot arm to generate perceptible vibrations in certain configurations. These vibrations do not occur when the robot manipulates lighter objects. More precisely, when the combined weight of the gripper and the object is close the robot's maximum payload, the built-in controller seems to generate motor commands (specifically in the last joints) that produce small but unwanted vibrations. These vibrations are not related to slippage but are still sensed by the tactile sensors when manipulating some specific objects. In an industrial context, this last class could have also been used to represent other types of disturbances, such as the ones generated by nearby-located machines.



a) Class 1: object-gripper slippage

b) Class 2: object-world slippage

c) Class 3: no relevant dynamic events

d) Class 4: external vibrations not related to slippage

Figure 3.4 Typical examples of raw dynamic data and spectrograms for each of the four considered classes

Fig. 3.4 depicts typical pairs of raw data and spectrograms for each of the four classes. It would be difficult to find a criteria for class discrimination by simply observing the raw dynamic data or the spectrograms; but we may still be able to find clues about some distinctive frequency patterns for each class by looking closely at their associated spectrograms. For example, the vibrations generated in the case of class 1 (object-gripper slippage - Fig. 3.4a) seem to typically dissipate energy mostly in a narrow, low frequency band ($\approx$ 0-200 Hz), while the energy dissipated by the vibrations from class 2 (object-world slippage - Fig. 3.4b) are spread across a wider band ($\approx$ 0-350 Hz). Furthermore, one can also note that the energy dissipated when no notable dynamic events are occurring (i.e. class 3 - Fig. 3.4c) is fairly homogeneously spread across the

complete 0-500 Hz bandwidth. Finally, when under the effect of external vibrations (class 4 - Fig. 3.4d), which in our case were due to the robot undesirably vibrating in specific conditions, the perceived vibrations seem to dissipate energy periodically, in a rather narrow band of low frequency similar to the one of class 1.

To encode the data in a way that the most relevant high level features are not-only preserved but also put on the front scene, we propose to use sparse coding (Olshausen & Field, 1996). To do so, we first collected 4728 unlabeled dynamic data, computed their spectrograms and used them to unsupervisedly learn a dictionary of features. These unlabeled data were randomly selected from data acquired during other robotic experiments, or generated on purpose from miscellaneous tactile events such as: tapping/rubbing/blowing air on the sensor, opening/closing the gripper, generating slippage on purpose and moving the robot with different accelerations while it was in contact (or not) with the environment.

When using sparse coding to learn a dictionary that captures high-level features of the input data, the resulting sparse encodings should ideally be able to provide a representation that disentangles the factors explaining the variations in the input data. At the same time, the sparse codes should also be invariant to some other uninformative factors that might also cause variations in the input (Bengio *et al.*, 2012). In our specific context, we want to have a dictionary ($\boldsymbol{D}$) that is invariant to small displacements of the box used to crop the dynamic data. For example, given $1,000$ points representing one second of tactile data (acquired at $1,000\ Hz$ while an unlabeled dynamic event was happening) and a data-cropping box of $500ms$, we ideally want our high-level features to be invariant, as much as possible, to small displacements of the $500ms$ cropping box that was used to select the data within this one second. Thus, to make the learned features more invariant to this factor, we artificially created a set of new data by duplicating each of the 4728 unlabeled data, and translating the duplicates, which was done by generating a small displacement of the cropping box used to get the original data. The cropping box was shifted by t_{shift} milliseconds:

$$t_{shift} = round(\frac{t_{Hamming} - t_{Overlap}}{2}), \tag{3.1}$$

where $t_{Hamming}$ and $t_{Overlap}$ are respectively the Hamming window size and the overlap (both in ms) used to compute the spectrograms. This shift value ensures the new data will not create a near-duplicate of an existing spectrogram. We then used these additional $4,728$ data along with the original data during the dictionary learning phase, in an effort to create more translated-cropping-box invariant features.

After the dictionary has been learned offline using easy-to-generate unlabeled data, it is then possible to use its contained features to reconstruct new spectrograms online. During online testing, the features from the dictionary that are used to reconstruct a spectrogram are sent to a simple linear SVM for event classification. Hence, another underlying topic that we explore in this paper is how well the specific features that were chosen from the dictionary to reconstruct a certain spectrogram allow its classification among the considered dynamic events.

3.3.1 Pre-processing algorithms

When robotic manipulation tasks are ongoing, it is often necessary to be able to quickly estimate whether a dynamic event is occurring and find out why it is happening. For example, when an object is slipping out of the gripper, it might be crucial for the robot to detect this situation in a short period of time. The robot must quickly recognize object slippage so it can take action, for example by adjusting its grasping force or by modifying its pose. For this reason, we considered only a $500ms$ period for detection and discrimination of the events' classes.

Whether it be to learn a dictionary of basis (top part of Fig. 3.3) or to classify the events during testing (bottom part of Fig. 3.3), we begin by computing a spectrogram out of a $500ms$ strip of dynamic tactile data acquired at $1,000\ Hz$. We experimented with various sizes of Hamming window (as will be discussed in more detail in section 3.5.1), but we always kept a constant overlap of 50%. In this work, we were able to discriminate the considered classes with the best overall classification accuracy when we used a Hamming window of $50ms$.

Then, to filter the obtained spectrograms, we inspired ourselves from how typical automatic speech recognition (ASR) algorithms filter data and applied a log Mel-filter bank (MFB) (Davis & Mer-

melstein, 1980) on all spectrogram's time bands. This transformation can be seen as an approximation of the human cochlea's behavior, which allows one to distinguish low frequencies more easily than high frequencies. Here, it is used to emphasize the fact that low frequencies may contain more relevant information for event discrimination than high frequencies. Also, as noted by Rispal (2014) in their work on texture recognition, when it comes to differentiating and detecting textures, the important region of the spectrograms to consider is the region specific to frequencies under 250 Hz. Furthermore, to detect slippage and to feel textures, humans are very efficient and they rely on specific mechanoreceptors to do so, particularly their Meissner and Pacinian corpuscles. These mechanoreceptors have a respective natural frequency of 30 Hz (Weerakkody *et al.*, 2007) and 250 Hz (Johnson, 2001), which is another reason why we decided to put more emphasis on the lower half of the frequency spectrum (below 250 Hz).

3.3.2 Sparse dynamic data encodings

Only few examples (e.g.: (Cockburn *et al.*, 2017; Liu *et al.*, 2016)) can be found in the literature where sparse coding has been used to encode tactile data and most of them are very recent . This is surprising given the possible biological foundations of this approach. Indeed, several studies have demonstrated that our human brain seems to encode at least some of our sensory information in a sparse manner (Baddeley *et al.*, 1997). In robotics, Rasouli *et al.* (2018) have emphasized the importance of using sparse encodings for sensory data. In their work, they were able to distinguish ten different textures with an accuracy of 92% using sparse tactile representations. Liu *et al.* (2016) have used sparse *static* tactile encodings for an object recognition task using a different sparse kernel for classification that takes into account, among other things, the spatial configuration of the tactile sensors. Here, we rather show how sparse *dynamic* data encodings can be used to efficiently discriminate important dynamic events. Although sparse tactile data encoding has not been profusely covered in the literature yet, sparse coding has on the other hand already been the subject of numerous works in other fields. Hence, here we only provide an overview of the general steps to sparsely encode a log MFB spectrogram. The main concept of sparse coding is to reconstruct a signal using a linear combination of a small amount of high

level features from a dictionary. These high level features are obtained by offline-learning a dictionary using unlabeled data.

The first step is to train a dictionary $\boldsymbol{D} := [\boldsymbol{d}_1, ..., \boldsymbol{d}_n] \in \mathbb{R}^{k \times n}$, of basis (features) $\boldsymbol{d}_k$, by solving

$$\min_{\boldsymbol{D},\boldsymbol{\alpha}} \sum_{i=1}^{m} \left(\left\| \boldsymbol{x}^{(i)} - \sum_{j=1}^{n} \boldsymbol{d}_j \alpha_j^{(i)} \right\|^2 + \beta \sum_{j=1}^{n} \left| \alpha_j^{(i)} \right| \right), \tag{3.2}$$

where $\boldsymbol{x}^{(1)}, ..., \boldsymbol{x}^{(m)} \in \mathbb{R}^k$ are the m segments of a spectrogram $\boldsymbol{X} \in \mathbb{R}^{k \times m}$ which has k power spectral density values. $\boldsymbol{\alpha}^{(1)}, ..., \boldsymbol{\alpha}^{(m)}$ are the sparse vectors that capture a small set of the most important high-level features of $\boldsymbol{X}$ from the dictionary $\boldsymbol{D}$ and β is an arbitrary non-sparsity penalty factor. Within the summation, the first term is the squared representation error penalty, which is defined as the squared difference between the real spectrogram segment $\boldsymbol{x}^{(i)}$ and its approximated reconstruction using dictionary elements. The second term penalizes non-sparsity representations and grows proportionally to the number of features that are used. Together, these two terms make sure to build a dictionary containing only the most useful high-level features that ensure the proper reconstruction of the unlabeled input spectrograms.

The dictionary learning problem (3.2) is a double optimization problem which is complex to solve and computationally intensive. Fortunately, Lee *et al.* (2007) show that this complex problem can be divided into two simpler-to-solve sub-problems that can also be computed iteratively. Fig. 3.5 depicts an example of such dictionary, and more specifically shows a part of the 140-basis dictionary that gave the best result during this work. Each vertical band corresponds to a high-level feature that was learned during offline dictionary training, while the color gives an indication of the computed power over the frequency bins. To the untrained eye, the basis shown in Fig. 3.5 could look like patches of noise, but it is rather a set of important high-level features. Indeed, for the sake of comparison, we also tried to build a dictionary made of pure Gaussian noise. While this worked to some extent, we found out that in this case, the sparsity level needed to be very low when compared to a real dictionary in order get similar good-quality reconstructions. This is logical, since a lot more random noise patches are needed

to reconstruct a spectrogram, in comparison to using patches from a properly-trained dictionary. Also, the classification success rates were obviously always significantly lower in this case and lead to a poor overall efficiency.

Once a dictionary $\boldsymbol{D}$ has been trained offline, it is then ready to be used online to sparsely encode new, never-seen data. This is done by solving the now-single optimization problem

$$\min_{\boldsymbol{\alpha}} \left\| \boldsymbol{x} - \sum_{j=1}^{n} \boldsymbol{d}_j \alpha_j \right\|^2 + \beta \sum_{j=1}^{n} \left| \alpha_j \right| . \tag{3.3}$$

The goal is to find sparse vectors $\boldsymbol{\alpha}^{(i)} \in \mathbb{R}^n$ that properly encode each $\boldsymbol{x}^{(i)}$, such that they could be reconstructed simply by computing:

$$\boldsymbol{x}^{(i)} \approx \sum_{j=1}^{n} \boldsymbol{d}_j \alpha_j^{(i)} \qquad\qquad i = 1, ..., m. \tag{3.4}$$

Since the sparse vectors $\boldsymbol{\alpha}^{(i)}$ indicates which dictionary basis are relevant to a certain spectrogram and how much they contribute to its reconstruction, it is what will be subsequently sent to a support vector machine (SVM) for classification.

3.4 Experiments

3.4.1 Setup description

To conduct the experimental part of this work, we used the capacitive-based tactile sensors presented in Fig. 3.2. All these sensor technologies were developed by our laboratory, the Control and Robotics Laboratory at École de technologie supérieure in Montreal, over the last few years and they are exhaustively presented in Rana & Duchaine (2013); Le *et al.* (2017). The "1st-gen" sensor can measure static pressure at a rate of 25 Hz with a resolution of 3×3 taxels, where each taxel covers an area of $4.2mm^2$. The "2nd-gen" sensor measures static pressure at a rate of 60 Hz with a resolution of 4×7 taxels, where each taxel covers $3.6mm^2$.

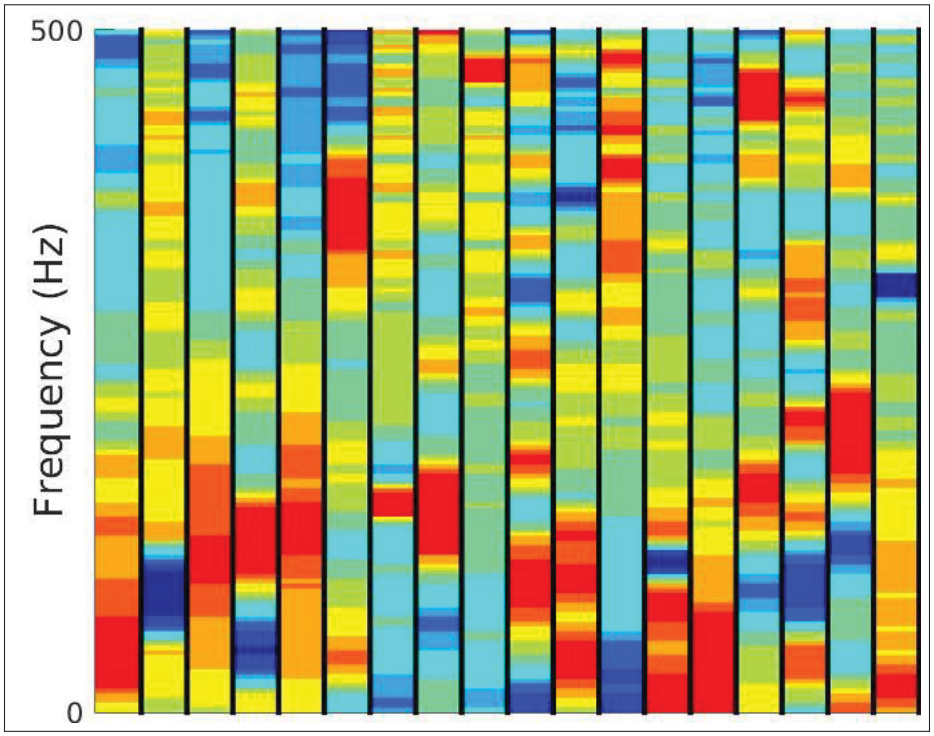


Figure 3.5 A sample of the 140-basis dictionary that gave the best results during this work

Inside both sensors, there is a two-layer microstructured dielectric that allows the detection of pressures as low as $10^{-4}N$ per taxel. In addition to measuring static pressure, the sensors also measure the *dynamic* pressure variations across its whole surface at a rate of $1,000$ Hz. Given this $1,000$ Hz sampling rate, the fast Fourier transform will then give power information only for frequencies less than 500 Hz, as specified by the Shannon-Nyquist sampling theorem. This dynamic sensing is achieved by amplifying the small current that comes out of (or goes in) the capacitor when pressure is applied on (or released from) the sensor with a transimpedance amplifier. This amplifier converts the small current to a voltage proportional to a gain that was chosen large. Thereby, the sensor is very sensitive to any kinds of vibrations, since these will generate small displacements of the capacitor's top electrode, which will in turn generate a current at the input of the transimpedance amplifier. In addition to this circuitry, the "2nd gen" sensor is also equipped with an 6-axis IMU, which provides another way of acquiring *dynamic*

data. Although the sensors generate both *static* and *dynamic* data, only the *dynamic* modality was used during this work.

The 1st-gen sensors were first integrated to a 3-Finger Adaptive Robot Gripper from Robotiq Inc which has a stroke of $155mm$ and a maximum payload of $7.5kg$. Although many 1st-gen sensors were integrated to the gripper (three per finger plus one on the palm), in this work we use only a single sensor located at the tip of the "middle" finger. Also, only the pinch mode of the gripper was used for all experiments. This sensorized gripper was mounted on a UR5 robot arm from Universal Robots, with a CB2 control box. This manipulator has 6 degrees of freedom and a maximum payload of 5 kg. The 2nd-gen sensors were only used to study the generalization potential of our method. They were mounted a different gripper, i.e. a 2F-85 gripper also from Robotiq Inc and also mounted on a different UR5 robot (although the latter being the same model, it is a different, distinct robot).

3.4.2 Data collection

The 1st-gen setup from section 3.4.1 was used to collect data by conducting a total of 972 experiments with 62 everyday-objects. Among these 972 experiments, there were 151 object-gripper slippage experiments (class 1), 289 object-world slippage experiments (class 2), 343 experiments with no dynamic events (class 3) and 189 external vibrations experiments (class 4).

Concerning the experiments related to class 1, we fixed the objects, turn by turn, in a clamp connected to a spring rigidly attached to a table. Then, we programmed the robot to pick up the object and to slowly move away from the spring's fixation point. As the robot moves the object away, the friction force gradually increases until it reaches a point where slippage will begin. Once we visually observed that slippage was occurring, we began recording a strip of dynamic data for each object until the gripper lost its hold on it. Regarding the experiments related to object-world slippage, we first programmed the robot to firmly grasp the object and then, we made it follow a precise trajectory that would generate slippage on either 1) a steel table (145 data) or 2) a wooden table (144 data). As with class 1, we recorded tactile dynamic data while

we visually noticed slippage was occurring. All of the 62 everyday objects have been used at least once during the experiments of class 1 and 2. For the third class, the experiments consisted in acquiring data while the robot was moving either holding nothing, holding a light object with a firm grasp or not moving at all. For class 4, we recorded tactile data while we made the robot generate undesired vibrations on purpose by putting a heavy object in the gripper and leaving the robot steady or making it move.

To control the robot and its attached gripper, as well as for recording the dynamic tactile data at 1000 Hz, we used ROS Kinetic running under Ubuntu 18.04 (Bionic Beaver). From the aforementioned 972 experiments' data, a relevant 500 ms portion of the dynamic tactile data was selected and was subsequently used to compute the spectrograms. The sparse coding algorithm that we employed is an adapted version of the MATLAB code available in Lee *et al.* (2007).

3.5 Results and analysis

3.5.1 Analysis of the hyperparameters' effects performance

As with most automatically classifying systems, in sparse coding there are many hyperparameters that influence the quality of the classification results. In Roberge *et al.* (2016), the entire set of hyperparameters was chosen through a trial-and-error method, resulting in a system that probably had a performance level below its full potential. In this section, we analyze the effects of some important parameters on the overall classification success rate. We seek to find rationales explaining the parameters' impact on the system's performance, such that a future implementer would better understand the effect of tuning these parameters. In particular, we show the impact of varying the following hyperparameters: the sparse penalty β, the spectrograms' frequency resolution, the Hamming window size, the number of allowed dictionary basis $\mathbf{d_i}$, and the mean sparsity level. We also deepen this topic by presenting a per-class analysis of the number of hits for each element of the dictionary. Logically, this analysis will allow to identify the set of parameter values that result in the highest success rate.

To study each of our selected parameters' effects on the system's performance, we varied one parameter at a time using a brute-force grid-search approach. Each of the aforementioned parameters was varied individually according to the process depicted by Fig. 3.3. A new dictionary was learned and the classification success rate was evaluated each time by taking the mean of the results obtained after 30 times a 10-fold cross validation was performed for each single combination of parameters. We evaluated a total of 1836 different combinations of parameters.

Since this grid-search scheme is both computationally expensive and time-consuming, we distributed the grid-search algorithm on 17 Linux-based Amazon EC2 instances that ran our compiled MATLAB's code over 20 days. The parameters' range and step values were set by hand by manual testing of different parameters prior to launching the algorithms. During these experiments, the overall classification accuracy performance indicator P that was chosen was

$$P = \frac{1}{4}\left(\frac{N_{SC1}}{N_{C1}} + \frac{N_{SC2}}{N_{C2}} + \frac{N_{SC3}}{N_{C3}} + \frac{N_{SC4}}{N_{C4}}\right) * 100\%, \tag{3.5}$$

where N_{SCi} is the number of correctly-classified data of label Ci and N_{Ci} is the total number of data with label Ci. This is thus a normalized classification success rate across all considered classes.

3.5.1.1 Effect of β

As mentioned in section 3.3, β corresponds to the importance given to getting data encodings that possess a high level of sparsity in the data encodings. Generally, when all the parameters are fixed except β, increasing β will result in more sparsely-encoded representations. Fig. 3.6 shows the average sparsity levels that was obtained for different values of β. Since it was known *a priori* that the sparsity level would not vary linearly with β, six unequally-distributed values of β were studied, which were $\beta = \{0, 0.1, 0.4, 4, 15, 40\}$. Fig. 3.7 shows the classification success rate for these values of β.

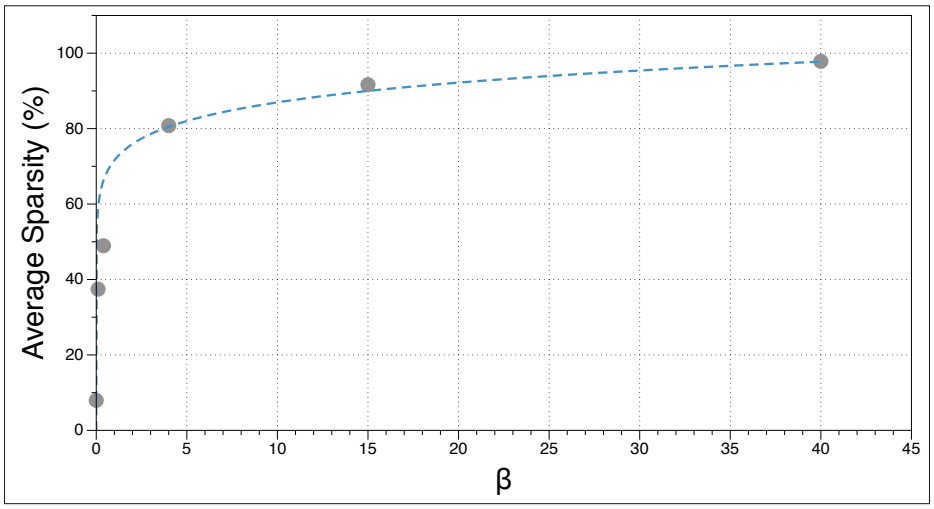


Figure 3.6 Sparsity level for different values of β

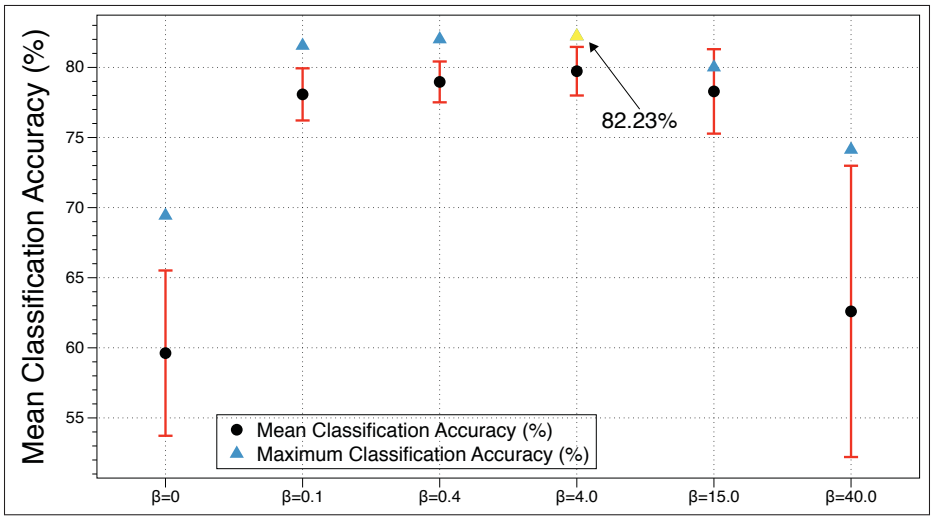


Figure 3.7 Mean classification success rate for different values of β, standard deviation is represented in red

It is apparent that mid-range values of β (between 0.1 and 15) might be better for obtaining good data representations. Here, the important point to notice is the trade-off to make between having not enough characteristic features of the considered events (high values of β) and having too many (low values of β), which will both complicate the classification process. It is also important to note that $\beta = 0$ corresponds to a special case where no regularization function is

used while solving the optimization problems of (3.2) and (3.3). This situation will obviously lead to representations that are not necessarily sparse and are thus less likely to capture any high-level information in the data. Conversely, $\beta = 40$ might lead to encodings that do not contains enough features to distinguish the event. This also explain why the standard deviation for these special cases are particularly high. The best result occurs when $\beta = 4$.

3.5.1.2 Effect of the number of basis (N_{Basis})

During the experiments, the number of dictionary elements (i.e. "n" in eq. 3.2 and eq. 3.3) was varied from 10 to 170, increasing the number in increments of 10. Fig. 3.8 shows the classification success rates for a number of basis between 20 and 170, since no coherent classification results were obtained with 10 basis. There is a slight increase of the success rates at the beginning, which indicates that a number of basis less than 70 might not have been sufficient for the dictionary to express a wide-enough variety of features to distinguish the events. On the other hand, having more than 120 basis seems to lead to poorer representations and increases confusion during classification. This might be explained by the increase of the number of basis without a consequent proportional increase of the amount of data to train the SVMs, such that there was an ever-growing amount of features to learn about a constant amount of data. The best result occurred when the number of basis is 140.

3.5.1.3 Effect of the frequency resolution

Here, the effect of changing the number of elements contained in each basis vector (i.e.: feature from the dictionary) is analyzed. This is literally the number of numerical values contained in each feature, which also corresponds to the number of bins used for computing each strip of spectrograms. The greater the number of elements, the greater the number of frequency bins over which the energy is computed for some time lapse. The results are shown in Fig. 3.9. The different numbers of elements per basis vector that were tested are $\{6^2, 8^2, 10^2, 12^2, 14^2, 16^2\}$. We used squared values because they enabled significantly-faster computation of our sparse coding algorithm.

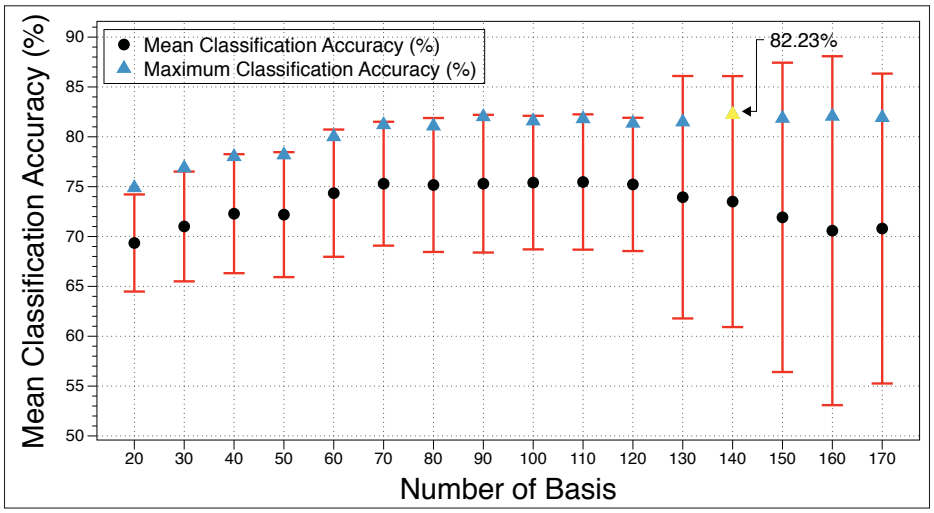


Figure 3.8 Classification results for different number of dictionary basis, standard deviation is represented in red

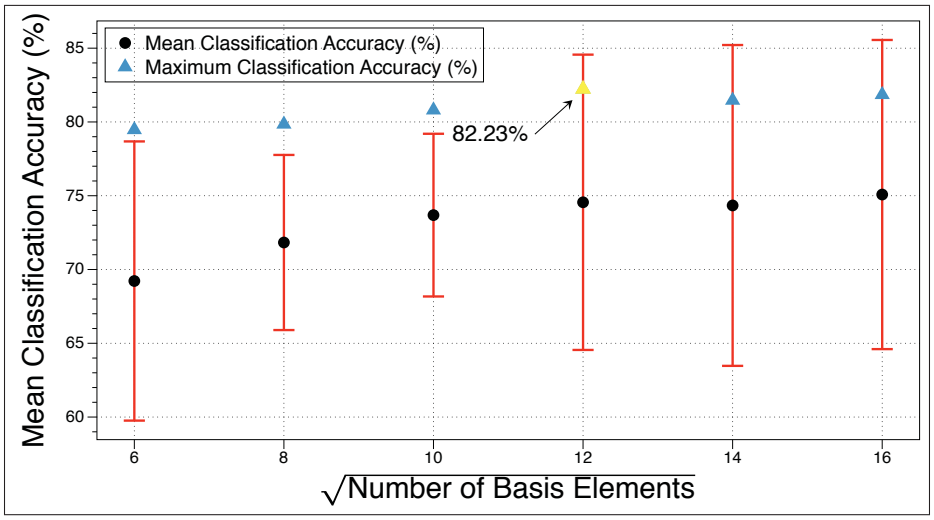


Figure 3.9 Classification success rates for the various numbers of elements per basis vector, standard deviation is represented in red.

The relation shows that if the frequency resolution is too coarse, features might become coarse too and lead to a non-optimal classification rate. However, the experiments also illustrate that after a certain degree of fineness, there might be no or very little value to continue to further increase the number of elements representing each feature. Correspondingly, here only 36 elements per basis vector were insufficient for providing good-quality representations—but once

100 elements were used, further increases in the number of elements did not seem to have much effect on the average classification success rate. The best classification results occur when the number of elements is 144 (i.e. 12^2).

3.5.1.4 Effect of the Hamming window size

During spectrogram computation, the Hamming window size is the period of time (within the $500ms$ window) over which the power spectral density is repeatedly computed. During the analysis, Hamming window sizes of $30ms$, $40ms$ and $50ms$ were tested, and an overlap of 50% was always chosen for each considered window size. Fig. 3.10 shows the classification results for these three window sizes.

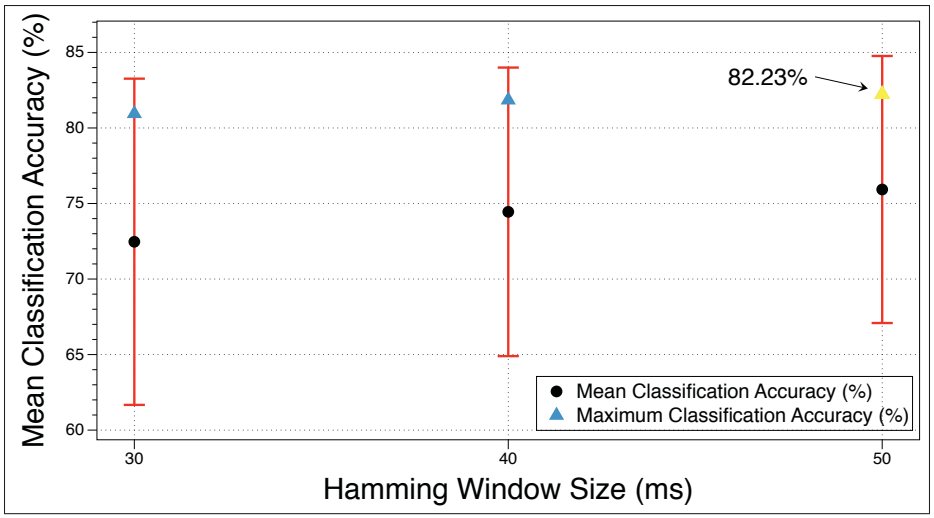


Figure 3.10 Classification accuracies for different Hamming window sizes, standard deviation is represented in red.

The larger is the Hamming window size, the smaller is the number of strips for a fixed $500ms$ spectrogram. The classification accuracy seems to exhibit a trend towards greater classification accuracy as the Hamming window size increases. More extensive testing would have been required to verify if this trend continues passed the highest considered value of $50ms$. However, there is a limit Hamming window size after which the classification rate will start decreasing.

For instance, the maximal Hamming window size of $500ms$, which is equivalent to computing the entire amount of energy contained in the recorded dynamic signal, was tested during this work and resulted in classification rates in the range of 60%-65%. That suggests there is again a trade-off to make between getting no granularity with how the energy is dissipated during the $500ms$ and having too much. The first case would fail to reveal the evolution of the energy dissipation during the event, which is likely to contain relevant information, while the second case would produce a large quantity of small strips that would not always capture relevant information.

3.5.1.5 Effect of sparsity

As with the conclusions drawn in Roberge *et al.* (2016), there seems to be a certain range of sparsity levels within which the classification results are better. Fig. 3.11 shows how the classification rate varies according to the level of sparsity found in the sparse vectors ($\boldsymbol{\alpha}^{(i)}$). When the sparsity is very low, the most important high-level features of the dynamic data are not very well highlighted. This complicates the classification problem since the most discriminative features of a considered event are likely to be mixed with other, non necessarily determinant features. On the other hand, when sparsity becomes very high, we start loosing important high-level features in favor of using a lower number of dictionary basis. At the limit when sparsity equals 100%, then absolutely no information remains and classification fails as the SVM is given no features at all. Here, the best classification results occur when the sparsity is $\approx 83.3\%$.

3.5.2 Analysis of dictionary elements usage per class

Many relevant observations come from analyzing the differences in the way each of the considered classes were encoded using specific dictionary elements. Here, the proportions of how much each dictionary element was used to represent the labeled data encodings for each class of events is analyzed.

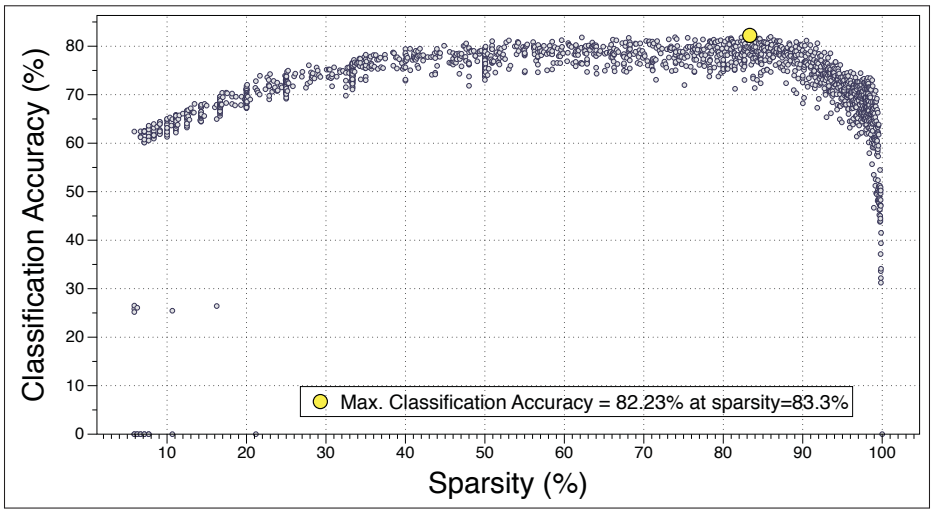


Figure 3.11 Classification accuracy at different sparsity levels

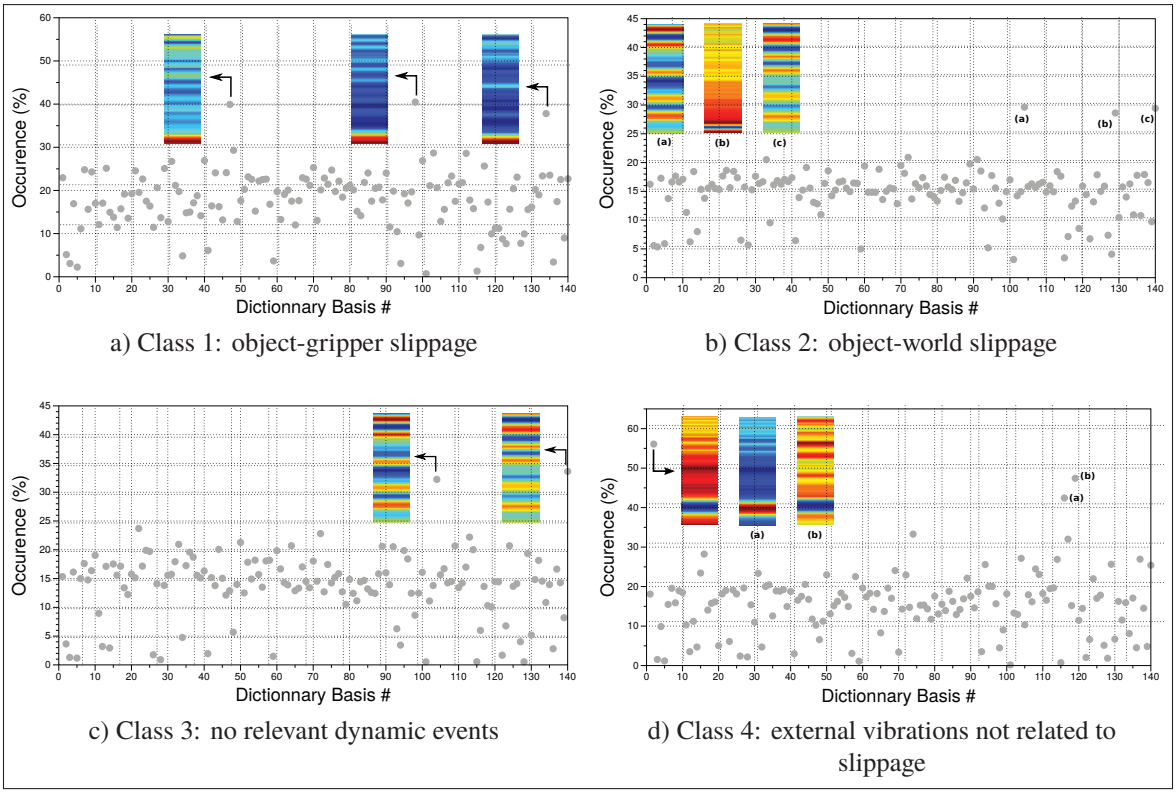


Figure 3.12 Dictionary elements' occurrence per class

Fig. 3.12 shows the percentage of usage of each dictionary element in all the data encodings per class. It also shows a visual representation of the basis that were used most often for each

class. As can be observed from Fig. 3.12a, the dictionary features that are the most often used to reconstruct the signals from class 1 is consistent with the typical spectrogram shown in Fig. 3.4a. Indeed, all three features seem to represent a concentration of energy within a thin region of the low-frequency band. Similarly, the most used features from class 3 seem to represent energy spread across various frequencies, as was noted in Fig. 3.4c. The most used features from class 4 also seem to be related to specific characteristics of this class' signals. Two out of three features represent the removal of energy in a specific low-frequency band, whereas the feature in the middle represents the addition of energy that also seems to be located in the same frequency band. This is coherent with the alternating energy dots of Fig. 3.4d. On the other hand, the case of class 2 is different because two of the most used features are visually similar as the ones most used to represent class 3. Because their energy is homogeneously spread across all frequencies, those two features intuitively seem to correspond better with class 3 than class 2. The fact that these two features were the most-used features for two classes could explain some of the confusion between these two specific classes. In overall, this analysis suggests that sparse encodings successfully captured the distinctive high-level features from the considered events automatically.

3.5.3 Results

In this section, we provide further details about the results obtained when using the best set of parameters from section 3.5.1. They are: $\beta = 4$, number of basis = 140, elements per basis vector = 144, and Hamming window size = $50ms$. These parameters produced an average classification performance (P) of $\approx 82.23\%$.

3.5.3.1 Pair-wise results

In Fig. 3.13, we analyzed the pairwise comparison of classification accuracies, in order to get an indicator of how easily each class of dynamic event could be distinguished from the others. To verify how well classes from each pair are distinguished from each other, we used 10-fold cross validations. In overall, the SVM classifier was able to properly distinguish each

pair of dynamic events at a very high success rate. The classifier had a slightly more difficult time processing data from classes 2 (object-world slippage) and 3 (null class), giving a lower normalized success rate of 76.2%. The apparent confusion between these two classes is most likely due to the fact that many of the 62 everyday objects that were used during the experiments had a soft-finished surface and only generated a very low level of vibrations when they were slipped across the support material's surface. Thus, the tactile sensors might not have perceived enough characteristic vibrations in some cases and confused these specific cases with the null class.

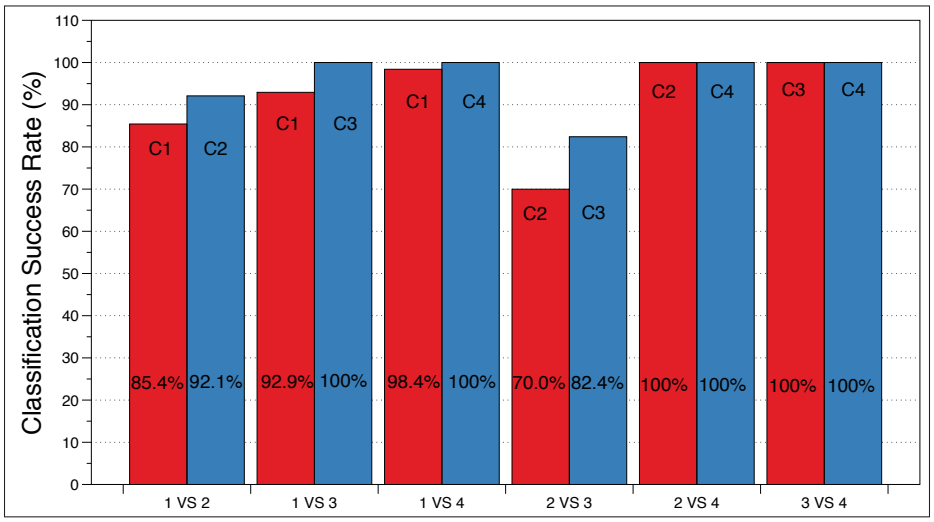


Figure 3.13 Pairwise comparison of classification accuracies

By contrast, class 1 (object-gripper slippage) was distinguished from class 2 (object-world slippage) with a normalized efficiency rate of 89.75%. Also, a spectacular result is when class 4 (external vibrations) is compared with any other considered dynamic events. Indeed, class 4 was always perfectly distinguished from all other events, indicating that the spectrograms related to this class contained very distinctive high-level features. This indeed seems to be true, as the robot-induced vibrations resemble to a set of periodical signals which are responsible for the periodically-appearing dots of energy at specific frequencies in most of this class' spectrograms.

3.5.3.2 Performance in different classification scenarios

To validate our approach when the four dynamic events are considered altogether, we used 10-fold cross-validations on all the 972 labeled spectrograms. Since SVMs are binary classifiers, but here the goal is rather to multi-classify each spectrograms into one of the four classes, we have turned to the well-known one-vs.-one classification technique. One-vs.-One is a reduction of the multi-classification problem and it consists of using several SVMs to evaluate each possible pair of classes (here, there are six possible pairs, thus six SVMs are required). Each SVM is responsible for voting for the class a given spectrogram most strongly correspond to and votes are compiled. At the end of the process, the class that had the greatest number of votes becomes the predicted class for the input spectrogram. One ambiguity with this procedure occurs when the maximum number of votes is even for more than one class. When this particular situation occurred, we used the SVMs' soft margins as an indication of the confidence level for each classification, and chose the even-voted class that had highest level of confidence. The obtained results are shown by the confusion matrix presented in Fig. 3.14.

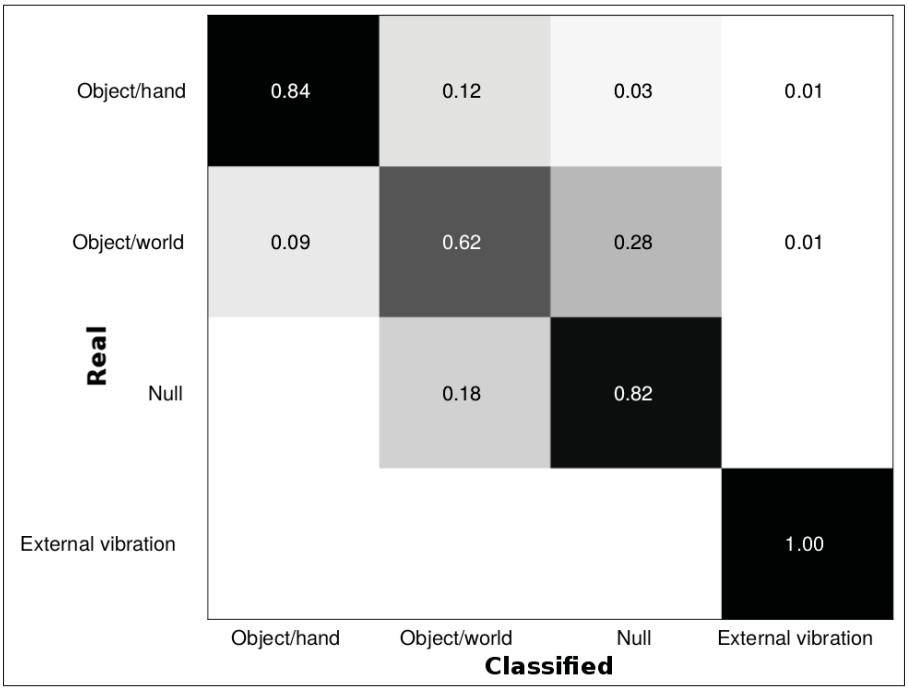


Figure 3.14 Confusion matrix

This confusion matrix shows that slippage was distinguished from the other considered dynamic events with a success rate of 84%, using only simple linear SVMs as classifiers. The fact that such a high success rate was obtained indicates how well the discriminative features were properly highlighted by sparse-encoding the data. The confusion matrix also shows some confusion between classes 2 and 3, which, as already noted in section 3.5.3.1, is most likely due to the fact that some of the 62 everyday objects didn't generate strong-enough vibrations when they were slipped across the support material to be perceived by the tactile sensors.

To extend our analysis with more possible classification scenarios, let's consider a set of input sparse vectors $\boldsymbol{\xi_i} \in C_i$, where C_i is the class $i \in \{1, 2, 3, 4\}$ such that, for example, $\boldsymbol{\xi_1} \in C_1$ means *the set of sparse vectors belonging to the object-gripper slippage class*. Let's also consider the notation $\boldsymbol{\xi_{1234}}$ which means *the set of input sparse vectors that belongs to classes 1, 2, 3 and 4 combined*, that is $\boldsymbol{\xi_{1234}} \in C_1 \cup C_2 \cup C_3 \cup C_4$. Finally, let D represent the set of considered classes from which the classifier needs to predict the membership of the input. Table 3.1 presents all classification rates for all possible additional scenarios.

Table 3.1 Classification accuracy for different scenarios

	D_{123}	D_{124}	D_{134}	D_{234}	D_{1234}
$\boldsymbol{\xi_1}$	84.87%	87.23%	90.89%	n/a	84.00%
$\boldsymbol{\xi_2}$	64.20%	91.83%	n/a	68.16%	62.5%
$\boldsymbol{\xi_3}$	82.46%	n/a	100%	82.46%	82.46%
$\boldsymbol{\xi_4}$	n/a	100.00%	100.00%	100.00%	100.00%
$\boldsymbol{\xi_{123}}$	80.69%	n/a	n/a	n/a	n/a
$\boldsymbol{\xi_{124}}$	n/a	92.66%	n/a	n/a	n/a
$\boldsymbol{\xi_{134}}$	n/a	n/a	97.00%	n/a	n/a
$\boldsymbol{\xi_{234}}$	n/a	n/a	n/a	86.65%	n/a
$\boldsymbol{\xi_{1234}}$	n/a	n/a	n/a	n/a	82.23%

Similarly to the aforementioned results, we used a one-vs.-one approach in conjunction with the SVMs' confidence level to perform all the multi-classification scenarios described here. Again, all the input sparse vectors corresponding to the fourth class (external vibrations) were always perfectly classified for all given scenarios. Another interesting point to mention is that

object-gripper slippage, which is often the most important event to detect, can be detected at a 90.89% success rate if one neglects the object-world slippage. Hence, when working in an application where object-world slippage will certainly not happen, it is possible to significantly increase the object-gripper classification success rate by more than 6%. When input vectors from all classes are considered, the overall classification success rate is 80.21%.

To quantify the beneficial effect of using sparse coding, we repeated the exact same experiments from sections 3.5.3.1 and 3.5.3.2, this time using raw, non-sparsely-encoded spectrograms. Without exceptions, all the obtained success rates were significantly lower in this context, most of the time being even less than half of those obtained with sparse coding. In comparison with sparsely-encoded spectrograms, the overall classification success rate obtained with raw spectrograms of all classes fell down to 39.81%, while the best and worse pair-wise success rates were respectively only 68.55% (class 4 VS. class 3) and 38.27% (class 2 VS. class 3). Of course, these rates could probably be improved if one is willing to use a more sophisticated classifier rather than a simple linear SVM. However, in general, it is important to highlight distinctive features from the input data as much as possible before sending the information to a given classifier. Sparse coding clearly accomplishes this task using an unsupervised algorithm and requires only unlabeled data, which is generally easy to get.

Table 3.2 Generalization analysis: success rates using new sensors

	2nd-Gen Capacitive Dynamic Tactile Sensor Data			**2nd-Gen Z-axis Accelerometer Data**		
	Case #1: New Dictionary, New SVM(s)	Case #2: Old Dictionary, New SVM(s)	Case #3: Old Dictionary, Old SVM(s)	Case #1: New Dictionary, New SVM(s)	Case #2: Old Dictionary, New SVM(s)	Case #3: Old Dictionary, Old SVM(s)
Pairwise						
Class 1 VS 2	87.36%	88.17%	79.29%	99.74%	99.43%	72.83%
Class 1 VS 3	95.01%	93.52%	86.63%	100.00%	100.00%	75.16%
Class 1 VS 4	97.11%	96.41%	85.75%	100.00%	100.00%	74.90%
Class 2 VS 3	69.06%	69.19%	65.38%	100.00%	100.00%	70.53%
Class 2 VS 4	72.99%	70.07%	68.72%	100.00%	100.00%	70.70%
Class 3 VS 4	70.47%	68.16%	62.44%	100.00%	100.00%	69.21%
One vs All						
Class 1	81.96%	79.23%	71.93%	99.57%	99.22%	70.48%
Class 2	54.31%	54.70%	45.26%	100.00%	100.00%	63.55%
Class 3	59.97%	56.42%	50.49%	100.00%	100.00%	65.26%
Class 4	68.24%	64.02%	56.19%	100.00%	100.00%	67.84%
All vs All	**65.93**%	**63.45**%	55.67%	**99.90**%	**99.82**%	66.76%

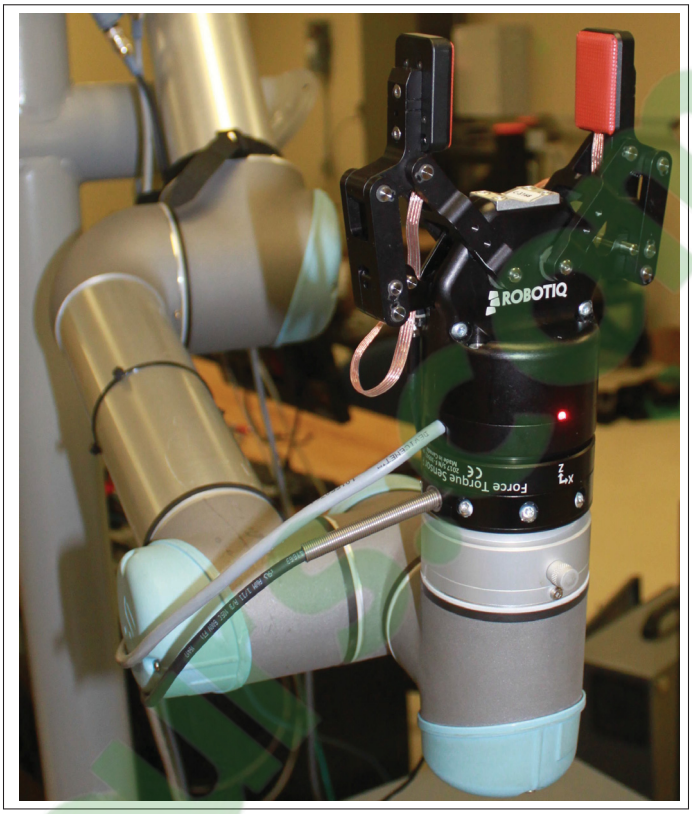


Figure 3.15 Close-up on the *new* setup used to study generalization: new tactile sensors (dynamic + accelerometer), gripper and UR5 robot

3.5.4 Generalization analysis

After having emphasized how well sparse coding could be put into contribution to highlight important features from the input data, we now analyze how well the proposed approach can generalize to other setups involving different gripper, robot and sensor technologies. To achieve this task, a completely different set of tactile sensors (i.e. the *2nd-gen* tactile sensors) was used, which was even mounted on a different kind of gripper and on a different UR5 robot. The sensor change will obviously modify the way vibrations from all classes are perceived, while the different robot and gripper might influence how vibrations from class 4 are generated and transmitted to the sensors. A close up view of this new and different setup is shown in Fig. 3.15. The two new tactile sensors (in orange) integrated to the gripper's fingertips are

described in Le *et al.* (2017). These latter also use capacitive sensing to acquire both static and dynamic data, but they are designed and built in a completely different fashion than the previous sensors. In addition to static and dynamic sensing, these new tactile sensors also include a built-in inertial measurement unit (i.e.: 3 accelerometers and 3 gyroscopes), allowing the use of the accelerometers as another mean for dynamic sensing.

These devices were used to collect data using the exact same procedure as the one described in section 3.4.2. For this analysis, a subset of 15 objects out of the whole set of 62 everyday objects was randomly chosen and was then used to collect a total of 652 dynamic and 652 accelerometer data, both divided as follows: 150 object-gripper slippage (class 1) data, 180 object-world slippage (class 2) data, 128 data extracted while no relevant dynamic event was happening (class 3) and 194 external vibrations (class 4) data. During the experiments, both tactile sensors were used to simultaneously record the capacitive dynamic data as well as the data from the accelerometers whose axis is normal to the sensor's surface (i.e. the z-axis). These data were then tested for three different cases. In the first case, unlabeled data generated by the new tactile sensors only were used to learn a new dictionary of features. Once this new dictionary was learned, new event-labeled data were acquired with the same sensors to train and test a new set of SVMs using 10-fold cross-validations. This allowed to establish baseline success rates by determining what level of performance could be achieved if the proposed approach from section 3.3 was completely reproduced using only the new sensors. In the second test case, the first dictionary that was learned with the 1st-gen tactile sensors (a part of which is shown in Fig. 3.5) was used. Here, the new data's spectrograms are encoded using only high-level features from the previous, totally-different, 1st-gen tactile sensors. The new data are only used to train/test new SVMs, but is never involved during the dictionary-learning phase. This allows to analyze how well the previous dictionary of features can generalize to other kinds of sensors. Finally, in the third case, the dictionary and the SVMs that were already trained with the 1st-gen sensors' data were used. Thus, the new data are solely used as a test set (absolutely no training was done using them). Now, this allows to analyze to what extent the SVM-and-dictionary combination can generalize to new sensors' data.

Results from these analysis are noted in Table 3.2. In overall, the new capacitive dynamic data seem to be less efficient for classifying the dynamic events than before, achieving a baseline success rate of only 65.93% when all classes are combined. This is most likely attributable to the significantly-reduced sensitivity of the new sensors' dynamic modality. Indeed, we noticed, by simply rubbing and tapping on the sensor, that the dynamic signal was way less reactive to any stimuli in comparison with the first set of sensors. This observation is also confirmed by comparing the design of the new 2nd-gen with the 1st-gen sensors. Indeed, while the 1st-gen sensor directly used its external surface made of a special barium titanate ($BaTiO_3$) doped silicon mix to measure vibrations (Rana & Duchaine, 2013), this 2nd-gen sensor rather has a neoprene covering that wraps around its external sensing layer, which attenuates the vibrations.

On the other hand, we noticed that the accelerometers were much more reactive than the capacitive dynamic modality of both considered tactile sensors. To our surprise, all the data were almost perfectly classified (99.90% of the time) using the proposed approach from this paper in conjunction with the accelerometers' data. Furthermore, to the best of our knowledge, these are the best classification success rates found in the literature for tactile event discrimination, considering additionally that they were obtained in the context of a classification problem involving no less than four distinct classes. Indeed, it was possible to almost-perfectly classify into one of four classes 652 accelerometer data, acquired on 15 different everyday objects which were not at all designed specifically for the experiments. To make sure that these impressive results were really as excellent as they seemed to be, further analysis were performed. More precisely, since classes #1 and #2 contain five to six data per objects which are randomly-shuffled before performing 10-fold cross-validations, it was important to make sure the SVMs didn't learn to classify their inputs using object-specific features, but rather features that are only related to the considered dynamic events. To do so, validations using k-fold cross-validations were done again, but this time by voluntarily isolating specific object-related-data from the training set and keeping them only for the test set. First, experiments were conducted where objects were isolated one-by-one (which corresponds to a non-random 15-fold cross-validation). Then, a second set of experiments were also carried out where data from all possible combinations of

2 to 5 objects within of the 15 objects were isolated. The lowest classification rate that was obtained during all these experiments was 96.0% which strongly suggests it was not the objects, but rather the events that were "memorized" by the SVMs. The small rate drop ($\approx 3\%$) might be simply explained by the fact that $1/3$ of the data was removed from the training set in the concerned experiment. On the other hand, many of the other aforementioned experiments ended with a perfect classification rate of 100.00%, confirming the authenticity of the results from Table 3.2. Both *old* and *new* dictionaries mentioned in Table 3.2, with the best hyperparameters from section 3.5 are publicly available[1].

Importantly, the results show that the dictionary of high-level features generalizes very well to other sensors, even to sensors like accelerometers that operate with a different technology. Indeed, building a dictionary from a single sensor seems to allow other sensors measuring the same underlying physical phenomenon to benefit from the learned features as well. In fact, when comparing any success rates between cases #1 and #2, one can see that they closely match most of the time. There might only be a very small deterioration of the rates when the dictionary was built using unlabeled data from another sensor than the one used for classification. However, when comparing cases #2 and #3, we clearly see that one could not use SVMs trained with another sensor without having to suffer an important success rate drop. Fundamentally, these results seem to indicate that dictionary features generalize to other sensors very well since they yield a similar success rate. However, which of these features are used and their respective level of activation will vary when using another kind of sensor, which creates the necessity to re-train the classifier(s).

3.6 Conclusion

While sparse coding is often used to efficiently encode data using a dictionary of high-level features, this paper has ehmphasized the fact that sparse encodings can also be used as reliable discriminants for classification. This indicates that the most useful features for reconstructing

[1] The dictionaries can be downloaded as MATLAB files at:
http://jproberge.net/SparseCoding/Dictionaries.zip

a signal might also contain information about the underlying nature of the said signal. Using these encodings, an almost-perfect classification (99.90% success rate) of four different dynamic events involving typical everyday objects was achieved using accelerometer data, surpassing our previous work's results (Roberge *et al.*, 2016) by a large margin. More importantly, it was shown that once a dictionary has been learned using a specific sensor, the same dictionary also generalizes very well to other sensors measuring the same physical variable, giving classification accuracies close to the ones that would have been obtained with an updated dictionary. Furthermore, it was shown how tuning the sparse coding hyperparameters affects the sparse representations' quality and the overall classification success rate, which gives clues about the important points to consider while implementing a similar technique. Using sparse encodings to represent sensory information can greatly reduce the quantity of data to handle, while still preserving the important features.

Since using the proposed approach makes detecting slippage more reliable and uses data acquired during only 500 milliseconds, it could be used to trigger re-grasping reflexes for robots in the future. We believe this work paves the way for developing robots that are more "tactile intelligent" and more capable of interpreting the tactile data it gets while they are interacting with their environment. Future work will study how much it is possible to minimize the amount of time required to properly classify dynamic events, in an effort to detect events as fast as possible.

CHAPTER 4

TACTILE-BASED OBJECT RECOGNITION USING A GRASP-CENTRIC EXPLORATION

Jean-Philippe Roberge[1], Louis L'Écuyer-Lapierre[1], Jennifer Kwiatkowski[1], Philippe Nadeau[1], Vincent Duchaine[1]

[1] Systems Engineering Department, École de Technologie Supérieure, 1100 Notre-Dame Ouest, Montréal, Québec, Canada H3C 1K3

Paper submitted to the International Journal of Advanced Robotic Systems (IJARS), August 2019.

4.1 Abstract

As humans, our grasping and manipulation skills are highly dependent on our ability to perceive tactile properties. Conversely, most of today's robotic operations still relies predominantly on visual feedback for identifying the objects that need to be grasped and manipulated. In this work, we rather study the problem of recognizing everyday objects based solely on their tactile attributes. This has a significant practical value, as it could allow object identification even when visual sensing is impossible, or assist vision in difficult contexts. Our method consists of acquiring multi-modal tactile sensing data during a quick and grasp-centric exploration phase, with minimal operational cost. Our algorithm was able to recognize objects from a considerably-large set of 50 general purpose items with an accuracy of 98.1%. Moreover, we show that it is possible to reliably identify a large proportion of these objects by only analyzing the deformation pattern that they undergo during the compression phase of their grasp. Further, we study our method's ability to learn relevant tactile properties to classify new objects. We also share our tactile sensing database that contains various sensor data acquired from more than 1,600 experiments, which was used for this work. Finally, we discuss the relative performance and role of each tactile modality for differentiating objects.

Keywords: Tactile Sensing, Object Recognition, Manipulation

4.2 Introduction

In the February 2007 edition of the Scientific American (Gates, 2007), Bill Gates anticipated that robotics would soon undergo a boom similar to the one the computer experienced in the 80s. The first materialization of this prediction has come in the form of a new, fast growing branch of robotics called collaborative robots. These robots, born out of a popular research trend that looks to harmoniously incorporate both robots and humans in the same workspace (Zinn *et al.*, 2004), have the advantage of being not only simpler to program, but also safer for humans. These particular technical characteristics have allowed collaborative robots, which appeared on the market in 2010, to quickly conquer a sector where automation was still not widespread, namely that of small and medium-sized enterprises (SMEs). This brilliant departure, as evidenced by the statistics from the International Federation of Robotics, represents just the tip of the iceberg as there still remains many opportunities for automation in these types of companies. Unfortunately, current collaborative robots do not yet have the capacity to perform the majority of tasks found in SMEs, where the production environment designed for humans is much less structured than that of the large production lines. One of the biggest technical bottlenecks that limits the potential massive integration of these robots in this sector is their poor grasping capabilities, which are well below the level of a humans. The commercial reality of SMEs, which often involves low-volume but high-diversity production, requires that any potential robots are able to reliably interact with a wide variety of parts and perform complex tasks such as assemblies. This is in considerable contrast to the situation currently prevailing in large factories, where robots are used mainly to tirelessly perform the same simple operation. Therefore, to maximize the potential of collaborative robots, it is imperative to raise their grasping and manipulation capabilities to a level that is more like that of humans.

In this new working paradigm, where robots have to grasp a multitude of different objects, being able to differentiate and recognize them becomes paramount. Furthermore, this ability could be useful for other tasks such as predicting an optimal grasp configuration, assessing grasp stability, or evaluating the object pose. Several researchers have sought to solve this problem by using artificial vision (Ekvall *et al.*, 2005; Bohg & Kragic, 2009).

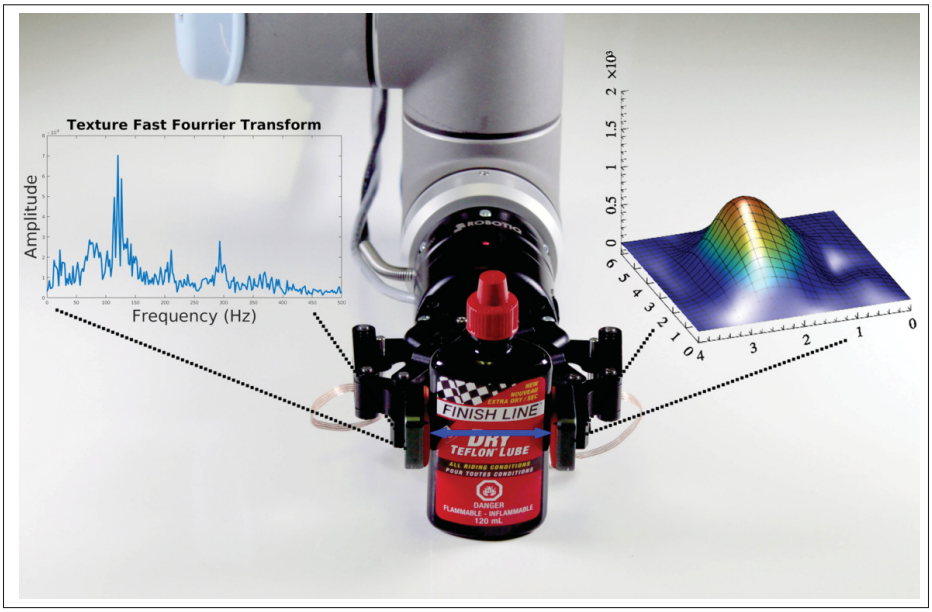


Figure 4.1 A gripper using a set of tactile sensors to extract relevant tactile properties

An example of an advanced vision-based robot is the well-known work presented in Kehoe *et al.* (2013) that uses the Google object recognition engine and a cloud-based planner to perform grasping. However, even the current state of the art vision systems suffer from technical limitations (e.g. lack of contrast, reflections, illumination, etc.), or task-related constraints (e.g. obstructed view by the gripper, cluttered environments, etc.) that introduce various uncertainties into vision-based operations. Moreover, while vision is good at mapping object geometry, it will generally fail to capture certain physical properties that pertain to an operation's outcome such as the object's stiffness, its surface texture, or its fragility. A particular example of this fact is shown in Fig. 4.2.

Rather than using artificial vision, the approach presented in this paper exploits tactile signals to recognize and differentiate objects. Recent works have demonstrated that the use of tactile signals allows the recognition of objects at a good success rate (Chen *et al.*, 2016a; Pezzementi & Hager, 2017; Bhattacharjee *et al.*, 2018; Lin *et al.*, 2019). Being able to rapidly recognize objects using touch would be a valuable skill in contexts such as assembly lines where a finite set of parts are to be picked up and identified, for example, from a shelf or from a bin where they are mixed.

Figure 4.2 Three visually-similar examples of different Starbucks mugs that have different tactile properties: The left mug is made of ceramic, the middle one is made of cardboard and the right one is made of plastic (All mugs could be either full or empty, which would also change their tactile properties)

This ability could potentially be used to assist or even replace vision in difficult contexts, for example when picking up parts in a cluttered and/or occluded environment, or where poor visibility prevails. In this vein, this work investigates the importance of different commonly-used tactile modalities for identifying typical everyday objects that were not designed or chosen specifically for the experiments. Particularly, the originality of this work lies in that it studies a variety of tactile feedbacks (proprioceptive, exteroceptive, static and dynamic) and discusses which one is better suited depending on the manipulation context. The contribution of each modality was evaluated on a large dataset encompassing over 50 objects. We deal with the advanced classification problem of automatically determining to which one of these 50 objects a specific, never-seen *tactile signature* belongs. Furthermore, while there are some examples of tactile-based object recognition in the literature, many of them rely on a more or less exhaustive tactile exploration phase. Thus, an additional contribution of this work is to propose a quick, grasp-centric exploration phase, that still allows the acquisition of some important tactile properties. Depending on the number and the nature of the objects to identify, we show that our

approach might only require tactile sensing data to be acquired during a typical object grasp, hence generating little-to-no additional operational cost.

This paper adheres to the following structure. First, the relevant literature on tactile sensing, especially in the context of object recognition, is introduced. Subsequently, the experimental approach implemented in this paper to study different modalities is outlined. Then, the experimental results are presented and analyzed. Finally, the significance of the obtained results as well as future work they inspire is discussed.

4.3 Related work

4.3.1 Tactile sensing in robotics

Integrating tactile sensors to enhance robotic operations is far from being a novel concept (R. Bajcsy & Khatib, 1984; Grimson & Lozano-Pérez, 1984). Typically, the purpose of a tactile sensor is to provide content-rich and task-relevant information about the tactile properties at the contact interfaces. Humans perform this feat using a dense array of biological mechanoreceptors, inherent to our sense of touch, to feel the different tactile modalities. The most predominant modalities used during manipulation tasks are normal pressure, vibration, and shear sensing (Johansson & Westling, 1984; Westling & Johansson, 1984). The development of a wide variety of tactile sensors just over the last years demonstrates the general effort towards incorporating these types of feedback into robots. Using different tactile sensing technologies (which include optical, resistive, magnetic, barometric, and capacitive-based mechanisms), there have been various attempts to build a sensor capable of measuring one or many of these modalities. General reviews of tactile sensing technologies can be found in Simpkins (2013); Cutkosky & Ulmen (2014); Cutkosky & Provancher (2016); Dahiya *et al.* (2010). In this work, we use the highly-sensitive multi-modal capacitive tactile sensor described in Le *et al.* (2017). It is further discussed in the *experimental setup* subsection. This particular sensor allows us to study the performance of different modalities for the purpose of recognizing objects given only tactile inputs, which is in fact a basic human ability (Klatzky *et al.*, 1985).

In recent years, tactile sensors have increased in popularity and are currently used in a wide variety of applications. These range from providing information about an object's pose (Bimbo *et al.*, 2016), an object's shape (Khasnobish *et al.*, 2014) or the contact force distribution (Sato *et al.*, 2017), to executing complex tasks such as automatic classification of dynamic events (Roberge *et al.*, 2016), grasp stability assessment (Cockbum *et al.*, 2017), or even the impedance-control of robots based on tactile feedback (Sato *et al.*, 2017). This work focuses on the task of recognizing objects by identifying their principal tactile characteristics.

4.3.2 Using tactile sensors for object identification

The concept of identifying objects using tactile sensing is a longstanding topic of research. For instance, in 1984, Grimson & Lozano-Pérez (1984) were among the firsts to study object identification using data from a tactile sensor. Twenty years after, Heidemann & Schopfer (2004) took advantage of the progress made in tactile sensing technologies over the years and used a time series of pressure profiles, obtained from a more advanced tactile sensor, to discriminate household objects. The approach, tested on seven objects, involved detecting a first contact, then exploring the surface of the object while maintaining a steady contact pressure. While the average classification success rate was good ($\approx 81\%$), the method involved a relatively complex exploration phase and was tested on a relatively small set of objects. On the other hand, Schneider *et al.* (2009) used pressure maps to discriminate 20 objects, but also required a significant exploration phase that involved touching the objects several times at different locations. While both of these last two mentioned works used *static data* (i.e. a pressure map acquired at a relatively low frequency), Chen *et al.* (2016a) rather measured vibrations at a high frequency (i.e.: *dynamic data*) and was able to accurately identify an object inside a container when it was shaken. This seems to indicate that *dynamic data* might also contains relevant information that could be used for the purpose of identifying an object. However, having to shake an object placed in a container might sometimes be problematic as it could potentially cause damage to the said object (e.g.: fragile objects) and has a significant time cost. Pezzementi & Hager (2017) proposed to use a set of two different handcrafted descriptors to identify objects, the first was

directly built from the tactile pressure map while the second was synthesized from the top-low frequencies in the Fast-Fourier Transform (FFT) of the same signal. While the two descriptors are different, they were handcrafted from the same source of information and were tested on five plastic letters, which might not reflect the wide variety of textures, rigidity, shapes and sizes, among other properties, that is normally found in household or industrial objects. Very recently, Lin *et al.* (2019) studied how a machine learning agent could learn to associate a tactile pressure map snapshot of a grasped object with its corresponding 2D image captured by a camera for object recognition. While the total number of considered objects was high (98), only a small subset of candidate objects (either 5 or 10) was presented at the time of classification with an accuracy ($\approx$ 24% to 40% for the first guess) that could make this solution impracticable in an industrial context.

While all of the aforementioned examples show interesting results, the literature still lacks examples where multiple tactile modalities (different sources of tactile sensing elements) are simultaneously employed to efficiently identify objects without having to rely on vision. Furthermore, most of the object recognition scenarios still involve a relatively time consuming and/or complex tactile exploration phase, with the accuracy capping or decreasing rapidly when the number of objects to recognize increases. Consequently, few approaches can be found where it is shown that the generated features can be used to learn the properties of a considerable quantity of arbitrary and non-specifically designed objects. In contrast with the literature, we propose a multi-modal approach utilizing two tactile sensors and a fast, grasp-oriented tactile exploration step. We quantify and compare the performances of different modalities with respect to their capacity to correctly recognize objects. We constructed a database[1] from 1600 experiments conducted on over 50 objects. The data was used to train different algorithms using either a single or a combination of the tactile modalities.

[1] This 2.5Gb database is shared as part of this submission here: www.dropbox.com/s/wumz72yetmp9zg6/Experimental%20Data.zip?dl=0

Figure 4.3 Some of the 50 objects used during the experiments

4.4 The approach

4.4.1 Experimental setup

The data for this work was acquired using the tactile sensor described in Le *et al.* (2017). It measures the normal pressures at 60 Hz using 28 capacitive sensors organized as a 4×7 taxel pressure image. A standard integrated inertial measurement unit (a MPU-9250 IMU, i.e.: three accelerometers, three gyroscopes. and a compass) provides proprioceptive feedback at 1kHz (which is particularly important for an under-actuated gripper) and a transimpedance amplifier measures the instantaneous capacitance variation at 1kHz. We refer to the 4×7 taxel matrix as

the *static data* and to both the instantaneous variation of capacitance and the accelerometer data as the *dynamic data*. A two-finger Robotiq© gripper was outfitted with two of these sensors, as shown in Fig. 4.1. The gripper was mounted on a UR10 robot from Universal Robots©. All the hardware was controlled and synchronized using ROS *Kinetic* on Ubuntu 16.04.

Using this setup, we performed a total of 1641 tactile exploration experiments on 50 objects (some are shown in Fig. 4.3) for an average of 33 experiments per object. 70% of this dataset was used for training (with 1/5 used for validation) and testing was done on the remaining 30%. We also conducted additional experiments (30 per object) on five never-seen objects and two previously-seen objects whose surface texture was altered on purpose to study which learned features would be selected in these cases.

4.4.2 The tactile exploration phase

Acquiring information using tactile sensors during a first-touch sequence should be simple, fast, and non-invasive while still providing enough information to infer rich features for the task at hand. Having these constraints in mind, we first began by investigating a three-step sequence (depicted in Fig. 4.1 and demonstrated in the accompanying video) to acquire data for each modality. Here we suppose the position of the object is fixed, or *a priori* known. Also, as this work does not study grasp planning, the pose of the gripper relative to the object was arbitrarily chosen at different locations on the objects prior to the robot sequence.

1. We move to the object's position and incrementally close the gripper until a first contact is detected. This is signified by the sum of the static taxels surpassing a given threshold value. We record the *opening of the gripper* right after the first contact to measure one of the object's dimensions;

2. Once slight contact has been established between the object and the tactile sensors, we initiate a "rubbing phase", by moving the robot's end effector downward by approximately one inch. During this phase, we continuously record the tactile *dynamic signals* consisting of the sensor's accelerometer feedback, measured from the axis normal to the sensor's

surface, and the instantaneous variation of the capacitance signal. The rationale behind this phase is to extract information that is a function of the object's surface properties, such as roughness and texture;

3. Finally, the gripper is commanded to close at half of its default speed and force to explore the object's stiffness. During learning, the object was squeezed five times during this phase for data collection (however it is only squeezed once during testing / actual object identification). For each squeeze we take four static data snapshots as shown in Fig. 4.5a and 4.5b. The idea is to capture the *deformation perceived at the fingertips*, essentially the evolution of the contact area and normal pressures during the compression.

4.4.3 The machine learning agents

To extract relevant high-level features from the tactile signals, an appropriate algorithm and/or machine learning technique must be selected. One of the primary challenges of any pattern recognition task is dealing with the nature of the data and its variability. This is particularity true in the case of tactile sensors, as Wan *et al.* (2016) demonstrated, tactile sensors often have a high variability in their feedback. In their work this resulted in unreliability when predicting grasp stability. Given these concerns, this section outlines our design of a robust agent capable of extracting pertinent information from the input signals.

4.4.3.1 The gripper opening position

As described in the *tactile exploration phase* section, as soon as the initial contact is detected between the object and the tactile sensors, the gripper opening position is recorded. This value is an integer that can vary between 0 (completely open) and 255 (completely closed) and is derived from the encoder attached to the motor that drives the gripper. Given a single object that is grasped multiple times at the same location, the gripper opening value (G) will slightly fluctuate between experiments due to different phenomena (e.g.: the gripper's repeatability, the tactile sensor's repeatability, permanent deformation in the object, etc.) Hence, here we assume

G is a random variable following an object-specific Gaussian distribution:

$$G_i \sim N_i\left(\mu_i, \sigma_i\right), \tag{4.1}$$

where N_i is the assumed Gaussian distribution of object i, with μ_i and σ_i being respectively the average and the standard deviation of the gripper opening at the time of initial contact with object i, where $i = \{1, 2, ..., 50\}$. During training, it is thus possible to characterize μ_i and σ_i for each object. After the training phase, we can then use these distributions to classify a gripper opening G by determining which class C is the most likely to correspond to the input G, yielding the highest probability for G:

$$C = \arg\max_i \left[\frac{1}{\sqrt{2\pi\sigma_i^2}} e^{-\frac{(G-\mu_i)^2}{2\sigma_i^2}} \right] \; i \in \{1, ..., 50\}. \tag{4.2}$$

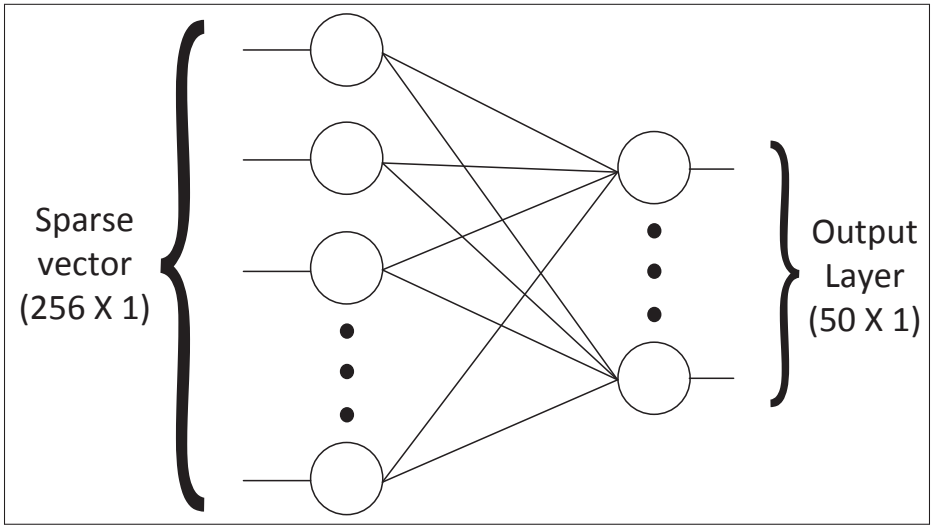


Figure 4.4 The two sparse vectors (128 elements each) are concatenated and linked to a simple fully-connected layer with 50 output nodes, where each node correspond to a specific object (here, the 128 elements are due to the 128 features contained in the dictionary, see Roberge *et al.* (2016))

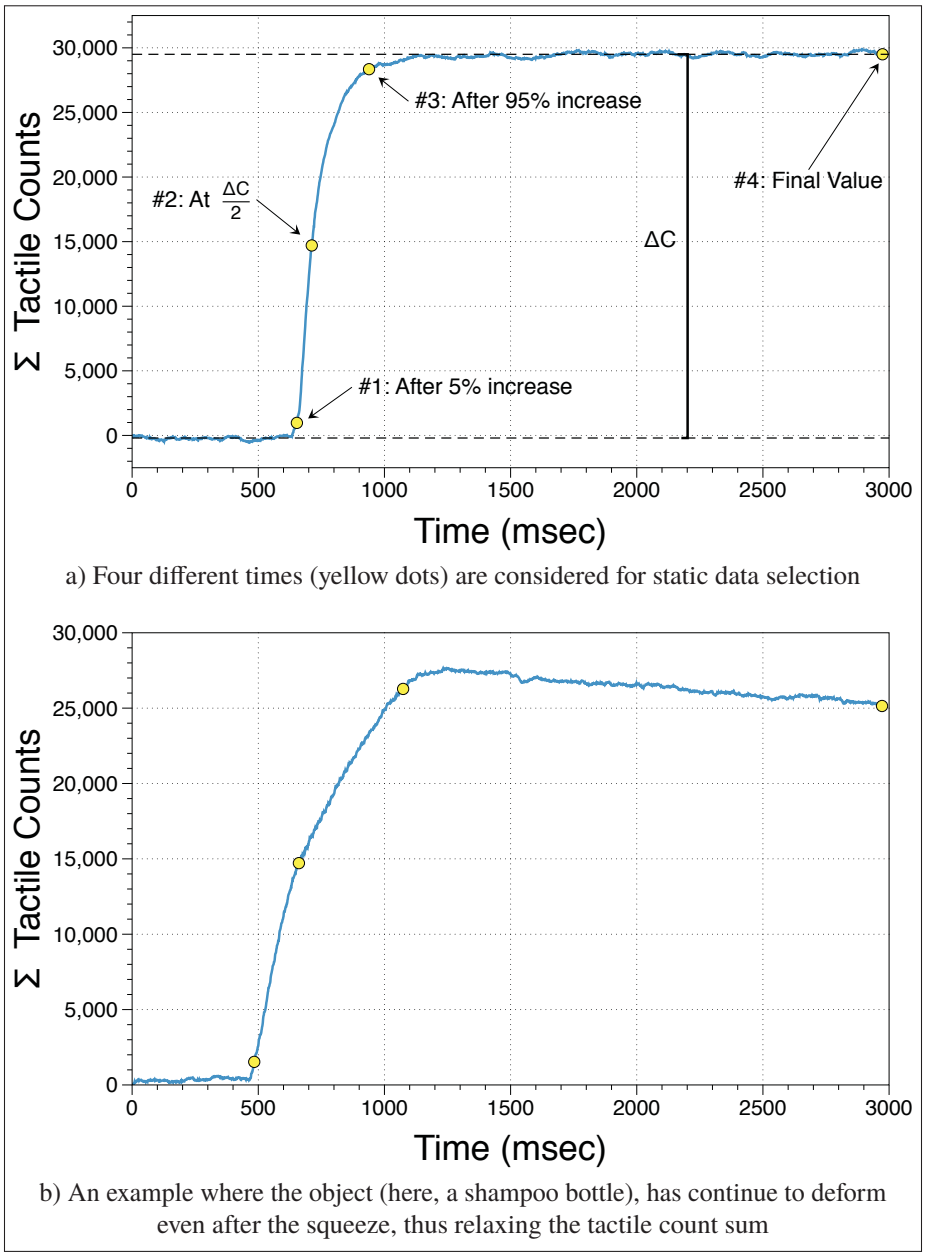
4.4.3.2 The dynamic data

The underlying hypothesis for this section is that the frequency component of a tactile sensor's dynamic signal will vary according to the object's surface properties. Hence, the first step for using this signal for classification was to convert it to the frequency domain using a regular Fast Fourier Transform (FFT).

In a previous work, performed with another tactile sensor (Roberge *et al.*, 2016), sparse coding was put into contribution to build a dictionary of 128 high-level features that was employed to successfully classify dynamic events at a high accuracy. The concept was to build a dictionary based on the most important features (from a signal reconstruction point of view) of a dynamic signal's FFT. However, the question of whether this dictionary could generalize to other sensors and/or tasks remained. Therefore, in this work we compared the classification success rate obtained by using a standard logistic regression on the FFT, to the success rate obtained using this previous dictionary of features. For the latter case, the dynamic signal's FFT was sparse encoded using the previously-learned dictionary and the resulting sparse vectors were fed to a simple fully-connected neuron layer, as shown in Fig. 4.4.

4.4.3.3 The perceived tactile deformation at the fingertips

The way an object deforms under compression is a tactile property influenced by both its geometry and stiffness, among others. The tactile data generated during this compression phase obviously contains a lot of meaningful information, since both the object's footprint on the fingertips and the normal pressure are evaluated at different moments and might be closely-related to each of the different considered objects. Also, it is important to note that our tactile sensors internally use a smooth dielectric providing it some degree of compliance too. Thus, even in the case of very rigid objects (e.g., rocks), the sensors will slightly adapt to the rigid object's shape as the grasping force increases, again revealing some object-specific features. It is the third and final tactile characteristic that we acquired during the aforementioned exploration phase.



a) Four different times (yellow dots) are considered for static data selection

b) An example where the object (here, a shampoo bottle), has continue to deform even after the squeeze, thus relaxing the tactile count sum

Figure 4.5 Times at which tactile data were acquired

During each squeeze, four tactile pressure maps (i.e. the concatenated *static* data of both tactile sensors) are taken at different times corresponding to: 1) when the sum of all the taxels reach 5% of maximum sum of the pressure distribution; 2) when the sum has reached half of the maximum value; 3) when the sum has reached 95% of the maximum value and 4) the final pressure maps of the compression phase acquired 3 seconds after the beginning of the squeezing

process and just before reopening the gripper (see Fig. 4.5). The first three snapshots encapsulate the evolution of the object's deformation (i.e. how the contact areas and the magnitudes of the normal pressures change) during the compression. Alternately, the last snapshot provides information about whether the internal forces within the grasp continue to change even after the gripper stopped moving. Fig. 4.5a is an example of a rigid object preserving its shape after the gripper has stopped its motion while Fig. 4.5b shows a softer object (a shampoo bottle containing a viscous fluid) that continues to change after the active compression is stopped. To benefit from both the object's footprint shape and the pressure maps information, the static data from each sensor were combined to create a side-by-side 7x8 taxel image at each of the four aforementioned instances. These pressure maps were then stacked to create a three-dimensional spatiotemporal input that was fed into the three-dimensional convolutional neural network shown in Fig. 4.6. The network extracts features that maintain their spatial and temporal relationships which helps classify objects based on their reaction to compression.

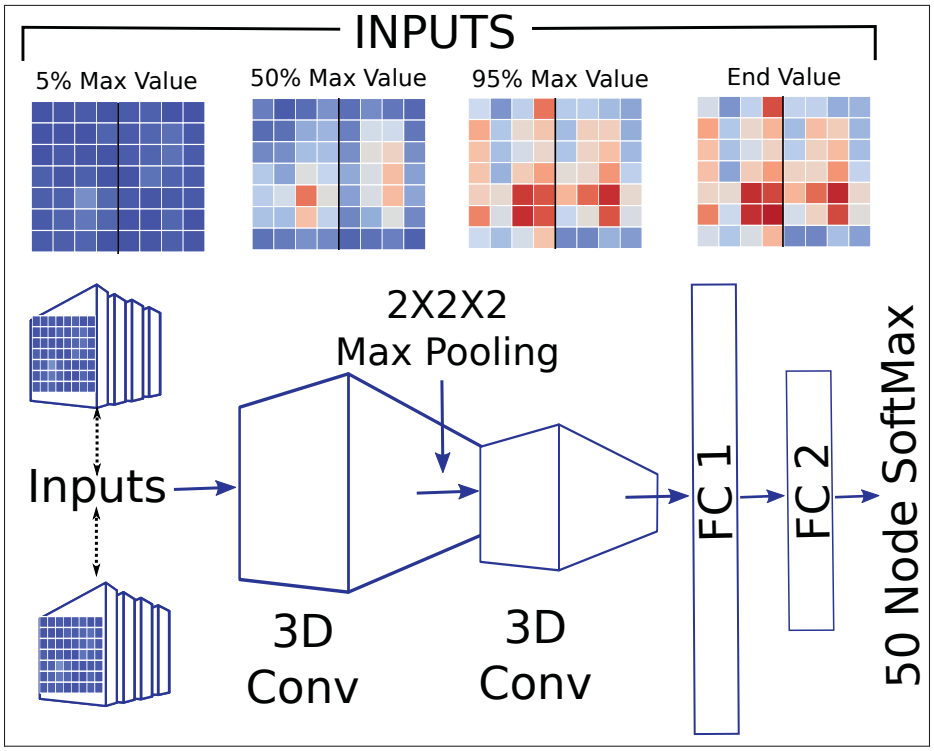


Figure 4.6 The 3D convolutional neural network used to process the perceived deformation at the fingertips

4.5 Experimental results and analyzes

4.5.1 The contribution of each modality

The tactile modalities described in the approach section were tested both independently and in conjunction with each other to evaluate their respective contributions to touch-based object recognition. To fuse different modalities together, we simply concatenated their respective outputs and connected that to two fully connected layers that were trained separately. The final results are presented in Table 4.1.

Table 4.1 Classification rate of each modality combination

Modalities	Classification Success Rate
Gripper Opening	67.1%
Dynamic Data^2	38.1%
Dynamic Data^3	48.6%
Perceived Object Deformation	**97.7**%
Opening + Dynamic^2	86.6%
Opening + Deformation	**98.1**%
Dynamic^2 + Deformation	**97.8**%
All together	**98.1**%

The top section of Table 4.1 shows the classification rate when only one modality is involved in the process. Surprisingly, by only squeezing the objects, we were able to achieve an impressive classification rate of 97.7% on 50 objects. This suggests that the information embedded in the progression of an object deformation contains a high level of unique features that can accurately characterize an item from the set. Thus, it appears that an object's deformation-related properties are more pertinent to object recognition than the other modalities.

To further analyze this impressive result, we used t-Distributed Stochastic Neighbor Embedding (t-SNE) (van der Maaten & Hinton, 2008) to visualize the compression data in a 2-D space. Fig. 4.7 shows the results obtained by applying this technique. Spectacularly, one can observe that the deformation data inherent to each object seems to gather in easily-distinguishable groups,

which seems to further confirm that each object from our dataset produced an almost-unique deformation pattern. This might be partly attributable to the complex but specific interactions between each single grasped object and our mildly-soft tactile sensor. Also, it might be relevant to note that there was a particularly large number of data acquired for this modality alone. Indeed, during the data collection process, we performed 5 squeezes per experiment (see the shared database), resulting in more than 4, 800 deformation patterns for the 50 objects.

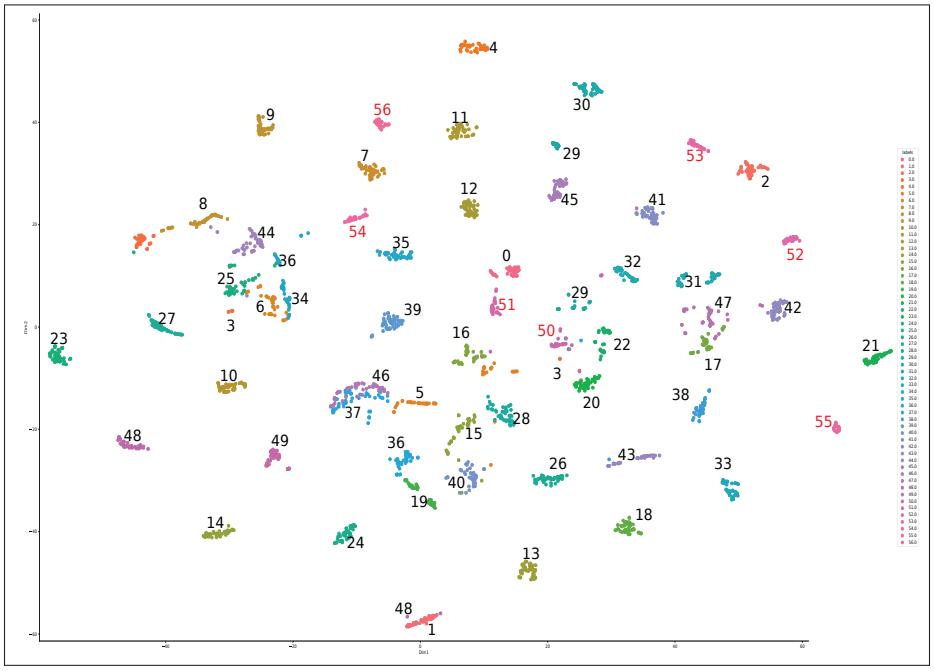


Figure 4.7 t-Distributed stochastic neighbor embedding (t-SNE)

The training dataset encompassed a wide range of diverse objects, thus one could have expected that the gripper opening position would produce a high classification rate. However, it was possible to achieve a 67% accuracy using this modality. This is likely the result of several factors such as groups of objects with similar dimensions, the variance in the tactile and gripper signals, and finally the discrete nature of the gripper command introducing gaps between the possible gripper positions.

The lowest classification rates were obtained using the dynamic modality. As previously mentioned, this modality is primarily dependent on the objects' surface properties. Here, this

modality has performed significantly less well than the others, which could be the symptom of several factors. For example, most of the objects either had a smooth surface or a similar surface roughness. This, coupled with the fact that our tactile sensors might not have a sensitive-enough dynamic modality could explain this lower classification success rate. On the other hand, keeping in mind that randomly classifying objects would lead to an expected success rate of 2% (=$1/50 \times 100\%$), it is evident that the dynamic data still contain a significant level of relevant information.

An interesting result to note is the important improvement in the classification rate when fusing the dynamic data with the gripper opening position as compared to using each of them independently. This would be useful in a case such as manipulating a fragile object, where the invasive squeezing phase is more likely to damage the objects. In this scenario, it would still be possible to achieve a respectable classification rate of 86.6%. The other combinations of modalities only slightly improve upon the results from the perceived deformation alone.

4.5.2 Sources of confusion

The classification results show that our machine learning agents can classify most of the objects at high accuracy, however there are some rare exceptions. Fig. 4.8 depicts the sources of confusion in each class. As expected, most of the tested objects are accurately classified, except in one case where there were 17 misclassifications which is significantly higher than the average of the other classes. This confusion happens between the can of peas and the coconut water bottle, as shown in Fig. 4.9. While this high number of misclassifications between these two specific objects seemed surprising at first, touching the objects ourselves and removing the thin paper label from the can of peas gave us a strong probable cause for this higher undesirable rate. Indeed, both objects are very rigid and do not deform much under pressure, they have a similar diameter (which is measured during grasping), but most of all: removing the label revealed they also seem to have an almost-identical corrugated surface. This corrugated surface is likely to create a similar dynamic signal when rubbed, since the grooves alternate in the same way.

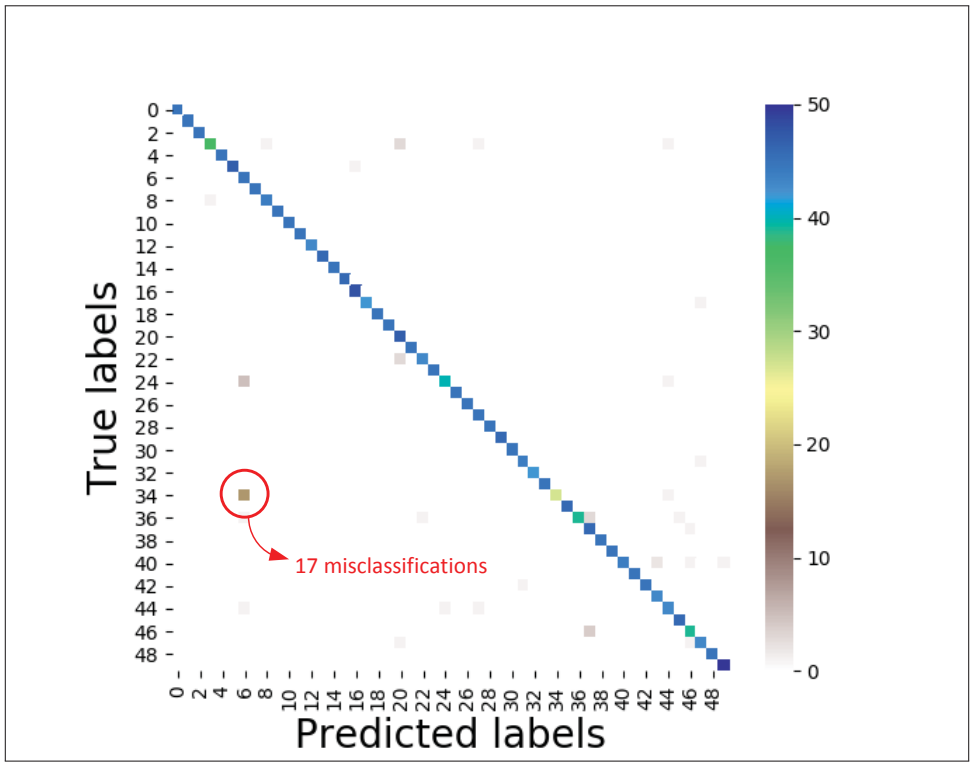


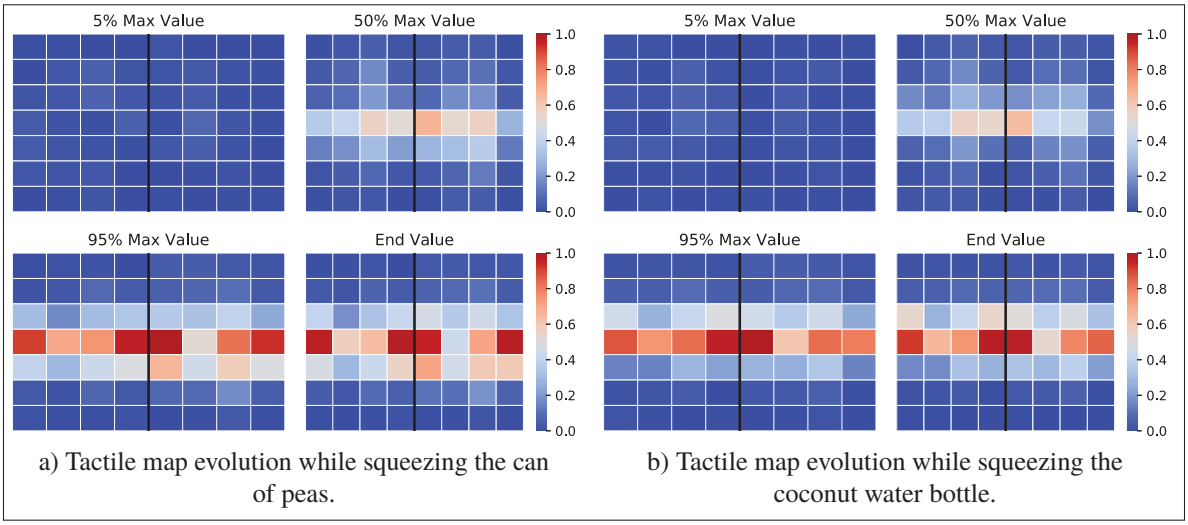
Figure 4.8 Confusion matrix when all modalities are used in synergy

Most importantly, depending on where they are touched, the two objects are very likely to produce a similar footprint and deformation pattern at the fingertips. While such misclassifications are undesirable, these two objects are very similar from a tactile sensing perspective. Fig. 4.10 shows their similar deformation behaviors. While this makes them difficult to differentiate, one could argue that, given their similar tactile signatures, both objects could be grasped using a similar strategy. There are analogous problematic cases in vision-based object recognition where two objects are visually very similar but have vastly different tactile properties, for example the three coffee cups in Fig. 4.2. However, in these latter cases, the objects would often need to be grasped using distinct strategies: for instance, a ceramic cup is very rigid but slippery, while a cardboard cup has a higher friction coefficient but is very deformable and cannot sustain the same level of normal grasping pressure.

Figure 4.9 Two objects that were confused during the evaluation made by the combined machine learning agents: a can of peas (shown on the left, its label was removed after the experiments) and a coconut water bottle – Conversely to Fig. 4.2, these visually-different objects are however similar from a tactile sensing perspective: they are both very rigid, they both have a smooth but corrugated surface, they have a similar radius (at the location where they were touched during the exploration phase) and a perceived deformation pattern (evolution of the contact area over time), when compressed, that looks alike

4.5.3 Never-seen objects and property inference

To investigate on how well the previously-described machine learning agents would individually deal with novel objects, we experimented on both never-seen objects and modified already-seen objects.



a) Tactile map evolution while squeezing the can of peas.

b) Tactile map evolution while squeezing the coconut water bottle.

Figure 4.10 The evolution of tactile maps during the squeezing phase for two confused objects

Table 4.2 shows how these objects are classified using a single modality at a time, with the two highest scoring classes being reported.

Table 4.2 Classification of never seen objects

Unseen Objects	**Modality**	**Classification**
Wood (small)	Static	60.0% (Wood), 20.0% (Stone)
Wood (small)	Dynamic	68.4% (Wood), 20.0% (Spoon)
Wood + Net	Static	63.3% (Wood), 23.3% (Fabric)
Wood + Net	Dynamic	80.8% (Net), 6.7% (Fabric)
Alu + S. P.	Static	90.0% (Board), 6.7% (Alu)
Alu + S. P.	Dynamic	89.6% (S.P.) 5.6%($Case$)
Perfume	Static	71.0% (Wood log), 25.8% (DVD)
Perfume	Dynamic	59.2% (DVD), 20.0% (Wood)
Cough Syrup	Static	100.0% (Salt)
Cough Syrup	Dynamic	47.2% (Container), 33.8% (Salt)

The first object that was presented was a wood block with significantly-smaller dimensions then the one used in our previous experiments. The contact area was smaller, so was the perceived deformation at the fingertips. Indeed, this was a rather important modification since this wood block would only partially cover the tactile sensors, which was never the case with

the wood block used in the previous experiments. Interestingly enough, that didn't stop the perceived deformation network from properly classifying this object most of the time (60.0%). Unsurprisingly, it was also properly classified by the dynamic data classifier, since the surface properties had remained the same.

The "Wood + Net" object is simply the original wood block, but this time covered with a thin layer of a net (from a previously seen object: see the *orange net* from our shared database) so that only its texture would change. As expected, the squeezing process still allowed a good classification rate for this scenario (63.3%), and the dynamic classifier was able to detect the texture at a high success rate. In fact, the netting's texture is so particular that it is exceptionally well classified (80.8%). Similarly, we also covered the aluminum part (previously seen) with sandpaper (also a previously seen object: see the *sandpaper medium 100* object). The dynamic data classifier made good predictions (it was right 89.6% of the time, again due to the fact that sandpaper has a very particular texture. On the other hand, this time the static data classifier has assigned a wrong object 90% of the time. This misclassification might indicate that covering the aluminum part with sandpaper slightly changed its deformation pattern. Indeed, the "board" is a flat cutting board that has a texture similar to the sandpaper. While squeezing the aluminum + sandpaper part, it is possible that the static pressure map showed a deformation pattern that is this time closer to the one of the cutting board. Despite this wrong assignation, the interesting fact is that we can still infer that the squeezed object was flat and rigid. In fact, for the majority of the misclassified static cases, the object was still mapped to one that had similar tactile properties, implying that the result could still be useful in a grasp planning or manipulation scenario.

Finally, two new and never-seen objects (a perfume and a cough syrup bottle) were introduced to our classifiers. The first is a rigid and flat bottle made of glass with a wood frame. The second is a plastic bottle with medium rigidity. What is interesting here is that although these objects were never memorized by our machine learning agents, the way they are classified relates to their tactile properties. For example, using only the classification results to infer tactile properties, one could have guessed that the perfume is likely to be a flat and rigid object, since both the wood log (which is basically another flattened piece of wood with a different texture) and the

aluminum were both flat and rigid objects that didn't change their shape during compression. It is also expected to have a smooth texture like that of the DVD case (whose texture was effectively similar to glass). Similarly, the cough syrup bottle was mapped to the only other plastic bottles (i.e.: the plastic container and the pink salt) from our dataset. Again, although the cough syrup bottle was never memorized or shown to our machine learning agent, we can still take advantage of the large variety of tactile properties that seem to have been learned by the algorithms to infer tactile properties to this new item as well. Indeed, given the classification results for this object, one could have guessed without making a mistake that it is cylindrical with a shape very similar to that of the pink salt container with a comparable medium rigidity (given the 100% static correspondence). This new object's texture might also be similar to the soft plastic surface finish of both the pink salt container and the plastic container.

4.6 Conclusion

As originally hypothesized, tactile sensors can be used to measure important tactile modalities such as perceived deformation, texture, and geometry which in turn can be used to recognize a wide variety of objects. The most impressive result of our work was our ability to identify objects with an accuracy of 98.1% using a combination of all the measured tactile signals. By far, the most useful information to properly recognize the considered objects was the perceived deformation at the fingertips during compression. Using only this information, it was possible to correctly classify the objects 97.7% of the time. Given this very high level of reliability, we propose to potentially only rely on this information to identify an object by recording the tactile pressure maps during grasping (see the accompanying video) at no additional production cost. Grasping an object using a significant but non-excessive level of force (in our case, the gripper's force and speed setpoints were both set to 50%) seem to exert enough compression to capture important object-related properties. The fact that the capacitive tactile sensor we used had a slightly compliant dielectric might also have helped in some cases, since it could have been able to locally comply with the shape of rigid objects, for example.

The dynamic modality was the lowest performer of all considered modalities which is possibly explained by the fact that our database is composed of many objects with a smooth surface. Additionally, the tactile sensor's dynamic sensing modality might not have been sensitive enough, more experimentation would be required to determine the specific causes. Also, to assess geometry, we exclusively used the hand-gripper opening after first contact, that is a value ranging between 0 and 255 which unfortunately resulted in the object's dimensions being measured with a coarse resolution and precision. Furthermore, as in the real world, many objects in our database had a similar diameter, resulting in a relatively poor level of successful classification when relying exclusively on geometry.

Nevertheless, we demonstrated that the proposed algorithms could be used to infer probable tactile properties of novel objects by matching them to similar objects that have been previously seen. In further works, it would be interesting to explore how the features extracted by our models could be generalized to aid in grasp planning and object manipulation. For example, an object with a very coarse texture may not require the same grasping strategy as a very smooth one depending on the object's weight. Furthermore, when manipulating a fragile object, information about the item's texture can be used to grab the object in a stable manner with as little force as possible to avoid damaging it. The features that were learned during grasping could be helpful in these contexts and should be investigated in future works. Finally, the inclusion of other tactile modalities, such as shear stress, could help create even more meaningful models in the future. This specific modality could be integrated to a similar grasp-centric approach as the one presented in this work with potentially no additional time, for example by recording shear during the lifting phase of the object. Such an important set of tactile properties could be useful for a wide variety of robotic tasks.

CONCLUSION AND RECOMMENDATIONS

For robots to have the same manipulation capabilities as humans, they need tactile intelligence. Before they can get there, greater advances in tactile intelligence must be made. To that end, this thesis aimed to advance robots' ability to grasp fragile objects, identify dynamic events like object slippage, and distinguish between different objects.

In a first paper, we demonstrated that industrial grippers could be improved by integrating tactile sensors onto their fingertips and by covering the sensors with a gecko-inspired dry adhesive to manipulate fragile and/or deformable objects. In particular, we showed that in the presence of such adhesives, the typical Coulomb friction law—where friction is proportional to normal force—no longer applies. Instead, what best describes the Van Der Waals interaction force between the grasped object and the adhesive is an adhesion model that is strongly correlated with the contact area. This means that we have a better way to handle fragile objects without damaging them. Instead of simply applying greater force, we can use this adhesive (and greater contact area) to grasp them effectively.

We also patented a chevron pattern for the adhesive that compensates—at low pressure—for tangential force, which prevents the object from slipping, and also compensates for moments, so that the object does not rotate within the gripper's grasp. It is even possible to predict the level of tangential force compensation one can expect to obtain with such adhesives by using tactile sensors to quantify the normal stress and contact area.

In a second paper, we used sparse coding to identify tactile events like object slippage. More precisely, we showed that using sparse coding to encode the frequency content of the vibrations measured by tactile sensors not only compresses the data, but also highlights important high-level features that can subsequently be used to identify whether the vibration is the result of object-gripper slippage, object-surface slippage, or external vibrations. We also showed that this method works for three different types of sensors.

Sparse coding thus enables detection of critical dynamic events, such as an object slipping from the gripper's grasp. This method uses a minimum of features to discriminate between dynamic events. And once the sparse coding algorithm has learned a dictionary offline, only a small set of features is required to distinguish and detect the important dynamic events.

Finally, in a third paper, we investigated three ways to identify an object from among 50 non-fragile objects: measuring vibrations when the gripper moved across the surface of an object; measuring the width of the gripper opening during the grasp; and measuring object deformity on the tactile sensors.

We studied how these methods can help identify objects when used both independently and together. We found that while they were all able to identify objects, by far the most efficient modality was *perceived deformation at the fingertips*. This method alone was enough to identify the vast majority of the objects. Moreover, the benefits of the object-deformation method are that there is no exploration phase, and vision is not required. Objects are identified at the same time as they are grasped, so there is theoretically no increase in processing time or operating cost.

In future, tactile intelligence could unlock greater capabilities for robots. To that end, it is very likely that AI algorithms using tactile inputs will have to be employed. However, one important challenge that may arise is the need to gather a large quantity of tactile data to properly train such algorithms. Given the fact that it is often difficult to collect a large number of such data, simulation might become useful to collect at least part of them. While there is already an appreciable number of robot simulators, both open-source and commercial, almost none are able to simulate tactile sensors and tactile interactions with the environment. Developing more evolved simulators could thus accelerate further developments in tactile intelligence, for the benefit of the whole robotics community.

APPENDIX I

LIST OF CONTRIBUTIONS

During the completion of this PhD thesis, the author contributed to the following journal papers, conference papers, and patents.

Papers submitted to journals with reviewing committee:

1. Gruebele A., **Roberge J.P.**, Zerbe A., Ruotolo W., Myung Huh T. and Cutkosky M., *A Stretchable Capacitive Sensory Skin for Exploring Cluttered Environments*. Published in the *IEEE Robotics and Automation Letters (RA-Letters)*, vol. 5, no. 2, pp. 1750-1757, January 2020.

2. **Roberge J.P.**, L'Écuyer-Lapierre L., Nadeau P., Kwiatkowski J. and Duchaine, V., *Tactile-Based Object Recognition Using a Grasp-Approach Style Exploration*. Submitted to the *International Journal of Advanced Robotic Systems*, 2019. Submission reference: ARX-19-0386 (under review).

3. **Roberge, J.P.**, Wong, T. and Duchaine, V., *Identifying Important Robotic Events Using Sparse Tactile Data Representations*. Submitted to IEEE Transactions on Robotics (T-RO), 2019. Submission reference: 19-0417 (first round of revisions).

4. **Roberge J.P.**, Ruotolo W., Duchaine V. and Cutkosky M., *Improving Industrial Grippers With Adhesion-Controlled Friction*. Published in the IEEE Robotics and Automation Letters (RA-Letters), vol. 3, no. 2, pp. 1041-1048, April 2018.

5. Motamedi, M.R., **Roberge, J.P.** and Duchaine, V., *The Use of Vibrotactile Feedback to Restore Texture Recognition Capabilities, and the Effect of Subject Training*. Published in the IEEE Transactions on Neural Systems and Rehabilitation Engineering. vol. 25, no. 8, pp. 1230-1239, Aug. 2017.

6. Rana, A., **Roberge, J.P.** and Duchaine, V., *An Improved Soft Dielectric for a Highly Sensitive Capacitive Tactile Sensor*. Published in the IEEE Sensors Journal. 16(22):7853-7863, 2016.

Papers submitted to international conferences with reviewing committee

1. Kwiatkowski J., **Roberge J.P.**, Nadeau N. A., L'Écuyer-Lapierre L. and Duchaine V., *An Extrinsic Dexterity Approach to the IROS 2018 Fan Robotic Challenge*. Published in the IEEE International Conference on Intelligent Robots and Systems (IROS) 2018, Madrid, Spain, pp. 4139-4144, 2018.

2. Cockburn D., **Roberge J.P.** and Duchaine V., *Grasp Stability Assessment through Unsupervised Feature Learning of Tactile Images*. Published in the IEEE International Conference on Robotics and Automation (ICRA) 2017, Singapore, pp. 2238-2244, 2017.

3. Le T., Maslyczyk A., **Roberge J.P.** and Duchaine V., *A Highly Sensitive Multimodal Capacitive Tactile Sensor*. Published in the IEEE International Conference on Robotics and Automation (ICRA) 2017, Singapore, pp. 407-412, 2017.

4. **Roberge J.P.**, Rispal S., Wong T. and Duchaine V., *Unsupervised Feature Learning for Classifying Dynamic Tactile Events Using Sparse Coding*. Published in the IEEE International Conference on Robotics and Automation (ICRA) 2016, Stockholm, pp. 2675-2681, 2016.

5. Motamedi, M., Chossat, J.-B., **Roberge, J.P.**, Duchaine, V. (2016). *Haptic Feedback for Improved Robotic Arm Control During Simple Grasping, Slippage, and Contact Detection Tasks*. Published in the IEEE International Conference on Robotics and Automation (ICRA) 2016, pp. 4894-4900, 2016.

Patents

1. Cutkosky M., Gruebele A., **Roberge J.P.** and Zerbe A., Stretchable sensory skin for robotic manipulation. United States Patent and Trademark Office (USPTO), Application 62/965846, *Provisional* patent application filed on January 25, 2020.

2. Cutkosky M., Ruotolo W. and **Roberge J.P.**, Patterned and Instrumented Directional Adhesives for Enhanced Gripping with Industrial Manipulators. United States Patent and Trademark Office (USPTO), Application 16/192,188, *Non-provisional* patent application filed on November 15, 2018. Ref: patents.google.com/patent/US20190143532A1/en

BIBLIOGRAPHY

Ackerman, E. (2019, June, 5). Amazon Uses 800 Robots to Run This Warehouse [IEEE Spectrum]. Consulted at https://spectrum.ieee.org/automaton/robotics/industrial-robots/amazon-introduces-two-new-warehouse-robots.

Alfadhel, A., Khan, M. A., Cardoso de Freitas, S. & Kosel, J. (2016). Magnetic Tactile Sensor for Braille Reading. *IEEE Sensors Journal*, 16(24), 8700-8705. doi: 10.1109/JSEN.2016.2558599.

Alpert, S. Z. Y. C. J. (1995). Spectral Partitioning: The More Eigenvectors, The Better. *Design Automation, 1995. DAC '95. 32nd Conference on*, pp. 195-200. doi: 10.1109/DAC.1995.250089.

Amend, J. R., Brown, E., Rodenberg, N., Jaeger, H. M. & Lipson, H. (2012). A Positive Pressure Universal Gripper Based on the Jamming of Granular Material. *IEEE Transactions on Robotics*, 28(2), 341-350. doi: 10.1109/TRO.2011.2171093.

Aquilina, K., Barton, D. A. W. & Lepora, N. F. (2018). Principal Components of Touch. *CoRR*, abs/1805.08039, 1–8. Consulted at http://arxiv.org/abs/1805.08039.

Aukes, D. M., Heyneman, B., Ulmen, J., Stuart, H., Cutkosky, M. R., Kim, S., Garcia, P. & Edsinger, A. (2014). Design and testing of a selectively compliant underactuated hand. *The International Journal of Robotics Research*, 33(5), 721-735. doi: 10.1177/0278364913518997.

Baddeley, R., Abbott, L. F., Booth, M. C. A., Sengpiel, F., Freeman, T., Wakeman, E. A. & Rolls, E. T. (1997). Responses of Neurons in Primary and Inferior Temporal Visual Cortices to Natural Scenes. pp. 1775–1783.

Bartolozzi, C., Ros, P. M., Diotalevi, F., Jamali, N., Natale, L., Crepaldi, M. & Demarchi, D. (2017, Sep.). Event-driven encoding of off-the-shelf tactile sensors for compression and latency optimisation for robotic skin. *2017 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pp. 166-173. doi: 10.1109/IROS.2017.8202153.

Begej, S. (1988). Planar and finger-shaped optical tactile sensors for robotic applications. *IEEE Journal on Robotics and Automation*, 4(5), 472-484. doi: 10.1109/56.20431.

Bekiroglu, Y., Kragic, D. & Kyrki, V. (2010, Sept). Learning grasp stability based on tactile data and HMMs. *RO-MAN, 2010 IEEE*, pp. 132-137. doi: 10.1109/ROMAN.2010.5598659.

Bengio, Y., Courville, A. C. & Vincent, P. (2012). Unsupervised Feature Learning and Deep Learning: A Review and New Perspectives. *CoRR*, abs/1206.5538, 1–30. Consulted at http://arxiv.org/abs/1206.5538.

Bhattacharjee, T., Clever, H. M., Wade, J. & Kemp, C. C. (2018). Multimodal Tactile Perception of Objects in a Real Home. *IEEE Robotics and Automation Letters*, 3, 2523-2530.

Bimbo, J., Luo, S., Althoefer, K. & Liu, H. (2016). In-Hand Object Pose Estimation Using Covariance-Based Tactile To Geometry Matching. *IEEE Robotics and Automation Letters*, 1(1), 570-577. doi: 10.1109/LRA.2016.2517244.

Bohg, J. & Kragic, D. (2009). Grasping familiar objects using shape context. *Advanced Robotics, 2009. ICAR 2009. International Conference on*, pp. 1–6.

Brodoceanu, D., Bauer, C. T., Kroner, E., Arzt, E. & Kraus, T. (2016). Hierarchical bioinspired adhesive surfaces—a review. *Bioinspiration & Biomimetics*, 11(5), 051001. Consulted at http://stacks.iop.org/1748-3190/11/i=5/a=051001.

Bronkhorst, A. W. (2000). The Cocktail Party Phenomenon: A Review of Research on Speech Intelligibility in Multiple-Talker Conditions. *Acta Acustica united with Acustica*, 86(1), 117-128. Consulted at http://www.ingentaconnect.com/content/dav/aaua/2000/00000086/00000001/art00013.

Büscher, G. H., Kõiva, R., Schürmann, C., Haschke, R. & Ritter, H. J. (2015). Flexible and stretchable fabric-based tactile sensor. *Robotics and Autonomous Systems*, 63, 244 - 252. doi: https://doi.org/10.1016/j.robot.2014.09.007. Advances in Tactile Sensing and Touch-based Human Robot Interaction.

Cavallo, A., Maria, G. D., Natale, C. & Pirozzi, S. (2014). Slipping detection and avoidance based on Kalman filter. *Mechatronics*, 24(5), 489 - 499. doi: http://dx.doi.org/10.1016/j.mechatronics.2014.05.006.

Caviglia, S., Valle, M. & Bartolozzi, C. (2014, June). Asynchronous, event-driven readout of POSFET devices for tactile sensing. *2014 IEEE International Symposium on Circuits and Systems (ISCAS)*, pp. 2648-2651. doi: 10.1109/ISCAS.2014.6865717.

Charalambides, A. & Bergbreiter, S. (2013, Nov). All-elastomer in-plane MEMS capacitive tactile sensor for normal force detection. *SENSORS, 2013 IEEE*, pp. 1-4. doi: 10.1109/ICSENS.2013.6688461.

Chebotar, Y., Kroemer, O. & Peters, J. (2014, Sep.). Learning robot tactile sensing for object manipulation. *2014 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pp. 3368-3375. doi: 10.1109/IROS.2014.6943031.

Chen, C. L., Snyder, J. O. & Ramadge, P. J. (2016a, Dec). Learning to identify container contents through tactile vibration signatures. *2016 IEEE International Conference on Simulation, Modeling, and Programming for Autonomous Robots (SIMPAR)*, pp. 43-48. doi: 10.1109/SIMPAR.2016.7862373.

Chen, Y., Yu, M., Bruck, H. A. & Smela, E. (2016b). Stretchable touch-sensing skin over padding for co-robots. *Smart Materials and Structures*, 25(5), 055006. doi: 10.1088/0964-1726/25/5/055006.

Cockbum, D., Roberge, J. P., Le, T. H. L., Maslyczyk, A. & Duchaine, V. (2017, May). Grasp stability assessment through unsupervised feature learning of tactile images. *2017 IEEE International Conference on Robotics and Automation (ICRA)*, pp. 2238-2244. doi: 10.1109/ICRA.2017.7989257.

Corradi, T., Hall, P. & Iravani, P. (2014). Tactile Features: Recognising Touch Sensations with a Novel and Inexpensive Tactile Sensor. *Advances in Autonomous Robotics Systems*, pp. 163–172.

Cutkosky, M., Jourdain, J. & Wright, P. (1987). Skin materials for robotic fingers. *Robotics and Automation. Proceedings. 1987 IEEE International Conference on*, 4, 1649–1654.

Cutkosky, M. R. & Provancher, W. (2016). Force and Tactile Sensing. *Springer Handbook of Robotics*, 717–736. doi: 10.1007/978-3-319-32552-1_28.

Cutkosky, M. R. & Ulmen, J. (2014). Dynamic tactile sensing. In *The Human Hand as an Inspiration for Robot Hand Development* (pp. 389–403). Springer.

Dahiya, R., Metta, G., Valle, M. & Sandini, G. (2010). Tactile Sensing:From Humans to Humanoids. *Robotics, IEEE Transactions on*, 26(1), 1-20. doi: 10.1109/TRO.2009.2033627.

Dang, H. & Allen, P. K. (2014). Stable grasping under pose uncertainty using tactile feedback. *Auton. Robots*, 36(4), 309–330. doi: 10.1007/s10514-013-9355-y.

Dargahi, J. (2000). A piezoelectric tactile sensor with three sensing elements for robotic, endoscopic and prosthetic applications. *Sensors and Actuators A: Physical*, 80(1), 23 - 30. doi: https://doi.org/10.1016/S0924-4247(99)00295-2.

Davis, S. & Mermelstein, P. (1980). Comparison of parametric representations for monosyllabic word recognition in continuously spoken sentences. *Acoustics, Speech and Signal Processing, IEEE Transactions on*, 28(4), 357-366. doi: 10.1109/TASSP.1980.1163420.

Day, P., Eason, E. V., Esparza, N., Christensen, D. & Cutkosky, M. (2013). Microwedge Machining for the Manufacture of Directional Dry Adhesives. *Journal of Micro and Nano-Manufacturing*, 1(1), 011001.

Deng, H., Zhong, G., Li, X. & Nie, W. (2017). Slippage and Deformation Preventive Control of Bionic Prosthetic Hands. *IEEE/ASME Transactions on Mechatronics*, PP(99), 1-1. doi: 10.1109/TMECH.2016.2639553.

Devadasan, S., Sivakumar, V., Murugesh, R. & Shalij, P. (2012). *Lean and Agile Manufacturing: Theoretical, Practical and Research Futurities*. PHI Learning. Consulted at https://books.google.ca/books?id=ECC3D6dtvOcC.

Donlon, E., Dong, S., Liu, M., Li, J., Adelson, E. & Rodriguez, A. (2018). GelSlim: A High-Resolution, Compact, Robust, and Calibrated Tactile-sensing Finger.

Eason, E. V., Hawkes, E. W., Windheim, M., Christensen, D. L., Libby, T. & Cutkosky, M. R. (2015). Stress distribution and contact area measurements of a gecko toe using a high-resolution tactile sensor. *Bioinspiration & biomimetics*, 10(1), 016013. doi: 10.1088/1748-3190/10/1/016013.

Edwards, J., Lawry, J., Rossiter, J. & Melhuish, C. (2008). Extracting textural features from tactile sensors. *Bioinspiration & Biomimetics*, 3(3), 035002. doi: 10.1088/1748-3182/3/3/035002.

Eisenhaure, J. & D, S. K. (2014). A Review of the State of Dry Adhesives- Biomimetic Structures and the Alternative Designs They Inspire. *Micromachines*, 8(4), 051001.

Ekvall, S., Kragic, D. & Hoffmann, F. (2005). Object recognition and pose estimation using color cooccurrence histograms and geometric modeling. *Image and Vision Computing*, 23(11), 943–955.

Fraden, J. (2003). *Handbook of Modern Sensors: Physics, Designs, and Applications (Handbook of Modern Sensors)*. SpringerVerlag.

Gates, B. (2007). A robot in every home. *Scientific American*, 296(1), 58–65.

Georgia Institute of Technology,Carnegie Mellon University, R. T. C. U. o. P. U. o. S. C. S. U. U. o. C. U. o. W. M. I. o. T. (2013). *A Roadmap for U.S. Robotics: From Internet to Robotics*. Consulted at https://books.google.com/books?id=KPhQngEACAAJ.

Goeger, D., Ecker, N. & Woern, H. (2009, Feb). Tactile sensor and algorithm to detect slip in robot grasping processes. *Robotics and Biomimetics, 2008. ROBIO 2008. IEEE International Conference on*, pp. 1480-1485. doi: 10.1109/ROBIO.2009.4913219.

Grimson, W. E. L. & Lozano-Pérez, T. (1984). Model-Based Recognition and Localization from Sparse Range or Tactile Data. *The International Journal of Robotics Research*, 3(3), 3-35. doi: 10.1177/027836498400300301.

Gruebele, A., Roberge, J.-P., Zerbe, A., Ruotolo, W., Huh, T. M. & Cutkosky, M. (2020). A Stretchable Capacitive Sensory Skin for Exploring Cluttered Environments. *IEEE Robotics and Automation Letters*, 1, 1–8.

Guler, P., Bekiroglu, Y., Gratal, X., Pauwels, K. & Kragic, D. (2014, Sept). What's in the container? Classifying object contents from vision and touch. *Intelligent Robots and Systems (IROS 2014), 2014 IEEE/RSJ International Conference on*, pp. 3961-3968. doi: 10.1109/IROS.2014.6943119.

Harrington, P. (2012). *Machine Learning in Action*. USA: Manning Publications Co.

Hartigan, J. A. & Wong, M. A. (1979). Algorithm AS 136: A K-Means Clustering Algorithm. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, 28(1), 100–108. Consulted at http://www.jstor.org/stable/2346830.

Hartigan, J. A. (1975). *Clustering Algorithms* (ed. 99th). New York, NY, USA: John Wiley & Sons, Inc.

Hawkes, E. W., Christensen, D. L., Han, A. K., Jiang, H. & Cutkosky, M. R. (2015, May). Grasping without squeezing: Shear adhesion gripper with fibrillar thin film. *2015 IEEE International Conference on Robotics and Automation (ICRA)*, pp. 2305-2312. doi: 10.1109/ICRA.2015.7139505.

Heidemann, G. & Schopfer, M. (2004, April). Dynamic tactile sensing for object identification. *Robotics and Automation, 2004. Proceedings. ICRA '04. 2004 IEEE International Conference on*, 1, 813-818 Vol.1. doi: 10.1109/ROBOT.2004.1307249.

Henaff, M., Jarrett, K., Kavukcuoglu, K. & LeCun, Y. (2011). Unsupervised Learning of Sparse Features for Scalable Audio Classification. *ISMIR*, pp. 681-686.

Heyneman, B. & Cutkosky, M. R. (2012, Dec). Biologically inspired tactile classification of object-hand and object-world interactions. *Robotics and Biomimetics (ROBIO), 2012 IEEE International Conference on*, pp. 167-173. doi: 10.1109/ROBIO.2012.6490961.

Heyneman, B. & Cutkosky, M. (2013, Nov). Slip interface classification through tactile signal coherence. *Intelligent Robots and Systems (IROS), 2013 IEEE/RSJ International Conference on*, pp. 801-808. doi: 10.1109/IROS.2013.6696443.

Heyneman, B. & Cutkosky, M. R. (2016). Slip classification for dynamic tactile array sensors. *The International Journal of Robotics Research*, 35(4), 404-421. doi: 10.1177/0278364914564703.

Hirai, Y., Suzuki, Y., Tsuji, T. & Watanabe, T. (2018, April). Tough, bendable and stretchable tactile sensors array for covering robot surfaces. *2018 IEEE International Conference on Soft Robotics (RoboSoft)*, pp. 276-281. doi: 10.1109/ROBOSOFT.2018.8404932.

Holweg, E., Hoeve, H., Jongkind, W., Marconi, L., Melchiorri, C. & Bonivento, C. (1996, Apr). Slip detection by tactile sensors: algorithms and experimental results. *Robotics and Automation, 1996. Proceedings., 1996 IEEE International Conference on*, 4, 3234-3239

vol.4. doi: 10.1109/ROBOT.1996.509205.

Hutchings, B. L., Grahn, A. R. & Petersen, R. J. (1994, May). Multiple-layer cross-field ultrasonic tactile sensor. *Proceedings of the 1994 IEEE International Conference on Robotics and Automation*, pp. 2522-2528 vol.3.

Hyttinen, E., Kragic, D. & Detry, R. (2015, May). Learning the tactile signatures of prototypical object parts for robust part-based grasping of novel objects. *Robotics and Automation (ICRA), 2015 IEEE International Conference on*, pp. 4927-4932. doi: 10.1109/ICRA.2015.7139883.

Hyvarinen, A. & Oja, E. (2000). Independent Component Analysis: Algorithms and Applications. *Neural Networks*, 13(4-5), 411–430. doi: 10.1016/S0893-6080(00)00026-5.

Hérault, J. & Ans, B. (1984). Circuits neuronaux à synapses modifiables: Décodage de messages composites par apprentissage non supervisé. *Comptes Rendus de l'académie des sciences*, 525–528.

Ibbotson, M. (2018, August, 3). Associates and Alphabot Team Up to Make Walmart's Popular Grocery Pickup Service Even Better. Consulted at https://corporate.walmart.com/newsroom/2018/08/03/associates-and-alphabot-team-up-to-make-walmarts-popular-grocery-pickup-service-even-better.

Ihlefeld, J. F. (2019). Chapter 1 - Fundamentals of Ferroelectric and Piezoelectric Properties. In Schroeder, U., Hwang, C. S. & Funakubo, H. (Eds.), *Ferroelectricity in Doped Hafnium Oxide: Materials, Properties and Devices* (pp. 1 - 24). Woodhead Publishing. doi: https://doi.org/10.1016/B978-0-08-102430-0.00001-2.

Israelachvili, J. N. (2015). *Intermolecular and surface forces*. Academic press.

Iwamura, Y. (2009). Tactile Senses – Touch. In Binder, M. D., Hirokawa, N. & Windhorst, U. (Eds.), *Encyclopedia of Neuroscience* (pp. 4005–4009). Berlin, Heidelberg: Springer Berlin Heidelberg. doi: 10.1007/978-3-540-29678-2_5873.

Jadhav, S. P., Wolfe, J. & Feldman, D. E. (2009). Sparse temporal coding of elementary tactile features during active. *Nature Neuroscience*, 1, 792–800.

Jiang, H., Hawkes, E. W., Fuller, C., Estrada, M. A., Suresh, S. A., Abcouwer, N., Han, A. K., Wang, S., Ploch, C. J., Parness, A. et al. (2017). A robotic device using gecko-inspired adhesives can grasp and manipulate large objects in microgravity. *Science Robotics*, 2(7), eaan4545.

Johansson, R. S. & Westling, G. (1984). Roles of glabrous skin receptors and sensorimotor memory in automatic control of precision grip when lifting rougher or more slippery objects. *Experimental Brain Research*, 56(3), 550–564. doi: 10.1007/BF00237997.

Johansson, R. S. & Flanagan, J. R. (2009). Coding and use of tactile signals from the fingertips in object manipulation tasks. *Nature Reviews Neuroscience*, 10(5), 345–359.

Johnson, K. O. (2001). The roles and functions of cutaneous mechanoreceptors. *Current Opinion in Neurobiology*, 11(4), 455 - 461. doi: http://dx.doi.org/10.1016/S0959-4388(00)00234-8.

Johnson, M. K. & Adelson, E. H. (2009, June). Retrographic sensing for the measurement of surface texture and shape. *2009 IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1070-1077. doi: 10.1109/CVPR.2009.5206534.

Kappassov, Z., Corrales, J.-A. & Perdereau, V. (2015). Tactile sensing in dexterous robot hands — Review. *Robotics and Autonomous Systems*, 74, 195 - 220. doi: https://doi.org/10.1016/j.robot.2015.07.015.

Kawamura, T., Inaguma, N., Nejigane, K., Tani, K. & Yamada, H. (2013). Measurement of Slip, Force and Deformation Using Hybrid Tactile Sensor System for Robot Hand Gripping an Object. *International Journal of Advanced Robotic Systems*, 10(1), 83. doi: 10.5772/55476.

Kehoe, B., Matsukawa, A., Candido, S., Kuffner, J. & Goldberg, K. (2013). Cloud-based robot grasping with the google object recognition engine. *Robotics and Automation (ICRA), 2013 IEEE International Conference on*, pp. 4263–4270.

Khasnobish, A., Singh, G., Jati, A., Konar, A. & Tibarewala, D. N. (2014). Object-shape recognition and 3D reconstruction from tactile sensor images. *Medical & Biological Engineering & Computing*, 52(4), 353–362. doi: 10.1007/s11517-014-1142-1.

Kinoshita, G.-I., Aida, S. & Mori, M. (1975). A pattern classification by dynamic tactile sense information processing. *Pattern Recognition*, 7(4), 243 - 251. doi: https://doi.org/10.1016/0031-3203(75)90009-6.

Klatzky, R. L., Lederman, S. J. & Metzger, V. A. (1985). Identifying objects by touch: An "expert system". *Perception & Psychophysics*, 37(4), 299–302. doi: 10.3758/BF03211351.

Klemm, M., Haueisen, J. & Ivanova, G. (2009). Independent component analysis: comparison of algorithms for the investigation of surface electrical brain activity. *Medical & Biological Engineering & Computing*, 47(4), 413–423. doi: 10.1007/s11517-009-0452-1.

Kulasekera, A., Amarasinghe, R. & Priyadarshana, P. (2013, 12). Quantum Tunneling Composite (QTC) Based Tactile Sensor Array For Dynamic Pressure Distribution Measurement. doi: 10.1109/ICSensT.2013.6727794.

Kwiatkowski, J., Cockburn, D. & Duchaine, V. (2017, Sept). Grasp stability assessment through the fusion of proprioception and tactile signals using convolutional neural networks. *2017 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*,

pp. 286-292. doi: 10.1109/IROS.2017.8202170.

Kyberd, P. J., Evans, M. & te Winkel, S. (1998). An Intelligent Anthropomorphic Hand, with Automatic Grasp. *Robotica*, 16(5), 531–536. doi: 10.1017/S0263574798000691.

Lavatelli, A., Zanoni, A., Zappa, E. & Cigada, A. (2019). On the Design of Force Sensors Based on Frustrated Total Internal Reflection. *IEEE Transactions on Instrumentation and Measurement*, 68(10), 4065-4074. doi: 10.1109/TIM.2018.2885604.

Lawrence, C. (2020, January, 1). Amazon Picking Challenge - The Future of Robotics. Consulted at http://amazonpickingchallenge.org/.

Le, T., Maslyczyk, A., Roberge, J. & Duchaine, V. (2017, May). A highly sensitive multimodal capacitive tactile sensor. *2017 IEEE International Conference on Robotics and Automation (ICRA)*, pp. 407-412. doi: 10.1109/ICRA.2017.7989053.

Lee, H., Battle, A., Raina, R. & Ng, A. Y. (2007). Efficient sparse coding algorithms. In Schölkopf, B., Platt, J. & Hoffman, T. (Eds.), *Advances in Neural Information Processing Systems 19* (pp. 801–808). MIT Press. Consulted at http://papers.nips.cc/paper/2979-efficient-sparse-coding-algorithms.pdf.

Lee, J., Shin, K.-y., Cheong, O., Kim, J. & Jang, J. (2015). Highly Sensitive and Multifunctional Tactile Sensor Using Free-standing ZnO/PVDF Thin Film with Graphene Electrodes for Pressure and Temperature Monitoring. *Scientific reports*, 5, 7887. doi: 10.1038/srep07887.

Lee, K., Ikeda, T., Miyashita, T., Ishiguro, H. & Hagita, N. (2011, Dec). Separation of tactile information from multiple sources based on spatial ICA and time series clustering. *2011 IEEE/SICE International Symposium on System Integration (SII)*, pp. 791-796. doi: 10.1109/SII.2011.6147549.

Lert, J. (2020, January, 1). Alert Innovation: Introducing Alphabot. Consulted at https://www.alertinnovation.com/.

Levine, S., Pastor, P., Krizhevsky, A. & Quillen, D. (2016). Learning Hand-Eye Coordination for Robotic Grasping with Deep Learning and Large-Scale Data Collection. *CoRR*, abs/1603.02199, 1–12. Consulted at http://arxiv.org/abs/1603.02199.

Li, M., Bekiroglu, Y., Kragic, D. & Billard, A. (2014, Sept). Learning of grasp adaptation through experience and tactile sensing. *Intelligent Robots and Systems (IROS 2014), 2014 IEEE/RSJ International Conference on*, pp. 3339-3346. doi: 10.1109/IROS.2014.6943027.

Li, R., Platt, R., Yuan, W., ten Pas, A., Roscup, N., Srinivasan, M. A. & Adelson, E. (2014, Sep.). Localization and manipulation of small parts using GelSight tactile sensing. *2014 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pp. 3988-3993. doi: 10.1109/IROS.2014.6943123.

Li, W., Zhou, Y., Poh, N., Zhou, F. & Liao, Q. (2013). Feature Denoising Using Joint Sparse Representation for In-Car Speech Recognition. *Signal Processing Letters, IEEE*, 20(7), 681-684. doi: 10.1109/LSP.2013.2245894.

Lin, J., Calandra, R. & Levine, S. (2019). Learning to Identify Object Instances by Touch: Tactile Recognition via Multimodal Matching. *CoRR*, abs/1903.03591, 1137–1143. Consulted at http://arxiv.org/abs/1903.03591.

Liu, H., Guo, D. & Sun, F. (2016). Object Recognition Using Tactile Measurements: Kernel Sparse Coding Methods. *IEEE Transactions on Instrumentation and Measurement*, 65(3), 656-665. doi: 10.1109/TIM.2016.2514779.

Luo, S., Mou, W., Li, M., Althoefer, K. & Liu, H. (2014, Nov). Rotation and translation invariant object recognition with a tactile sensor. *SENSORS, 2014 IEEE*, pp. 1030-1033. doi: 10.1109/ICSENS.2014.6985179.

Madry, M., Bo, L., Kragic, D. & Fox, D. (2014, May). ST-HMP: Unsupervised Spatio-Temporal feature learning for tactile data. *Robotics and Automation (ICRA), 2014 IEEE International Conference on*, pp. 2262-2269. doi: 10.1109/ICRA.2014.6907172.

Maria, G. D., Falco, P., Natale, C. & Pirozzi, S. (2015, May). Integrated force/tactile sensing: The enabling technology for slipping detection and avoidance. *2015 IEEE International Conference on Robotics and Automation (ICRA)*, pp. 3883-3889. doi: 10.1109/ICRA.2015.7139740.

Meier, M., Walck, G., Haschke, R. & Ritter, H. J. (2016a, Oct). Distinguishing sliding from slipping during object pushing. *2016 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pp. 5579-5584. doi: 10.1109/IROS.2016.7759820.

Meier, M., Patzelt, F., Haschke, R. & Ritter, H. J. (2016b). Tactile Convolutional Networks for Online Slip and Rotation Detection. *ICANN (2)*, 9887(Lecture Notes in Computer Science), 12–19.

Melchiorri, C. (2000). Slip detection and control using tactile and force sensors. *Mechatronics, IEEE/ASME Transactions on*, 5(3), 235-243. doi: 10.1109/3516.868914.

Nakamura, K. & Shinoda, H. (2001). A Tactile Sensor Instantaneously Evaluating Friction Coefficients. In *Transducers '01 Eurosensors XV: The 11th International Conference on Solid-State Sensors and Actuators June 10 – 14, 2001 Munich, Germany* (pp. 1402–1405). Berlin, Heidelberg: Springer Berlin Heidelberg. doi: 10.1007/978-3-642-59497-7_331.

Ng, A. Y., Jordan, M. I. & Weiss, Y. (2001). On Spectral Clustering: Analysis and an algorithm. *ADVANCES IN NEURAL INFORMATION PROCESSING SYSTEMS*, pp. 849–856.

Olshausen, B. & Field, D. (1996). Emergence of Simple-Cell Receptive Field Properties by Learning a Sparse Code for Natural Images. *Nature*, 381, 607-609.

OpenAI, Andrychowicz, M., Baker, B., Chociej, M., Józefowicz, R., McGrew, B., Pachocki, J. W., Pachocki, J., Petron, A., Plappert, M., Powell, G., Ray, A., Schneider, J., Sidor, S., Tobin, J., Welinder, P., Weng, L. & Zaremba, W. (2018). Learning Dexterous In-Hand Manipulation. *CoRR*, abs/1808.00177, 1–27. Consulted at http://arxiv.org/abs/1808.00177.

Paul Peter Urone, R. H. (2012). *College Physics - Capacitors and Dielectric* (ed. 1). Houston: OpenStax.

Pezzementi, Z. & Hager, G. D. (2017). Tactile Object Recognition and Localization Using Spatially-Varying Appearance. *Robotics Research : The 15th International Symposium ISRR*, 201–217. doi: 10.1007/978-3-319-29363-9_12.

Poczos, B. & Tishirani, R. (2016). CMU - Lecture 22: Independant Component Analysis. [Online; accessed 10-April-2016], Consulted at https://www.cs.cmu.edu/~bapoczos/Classes/ML10715_2015Fall/ica.html.

Pritchard, T. & Alloway, K. (1999). *Medical Neuroscience*. Wiley. Consulted at https://books.google.com/books?id=m7Y80PcFHtsC.

R. Bajcsy,Christensen, H. I. & Khatib, O. (1984). What can we learn from one finger experiments? *Robotics Research : The First International Symposium ISRR*, 509–527.

Raina, R., Battle, A., Lee, H., Packer, B. & Ng, A. Y. (2007). Self-taught Learning: Transfer Learning from Unlabeled Data. *ICML '07: Proceedings of the 24th international conference on Machine learning*.

Rana, A., Roberge, J. & Duchaine, V. (2016). An Improved Soft Dielectric for a Highly Sensitive Capacitive Tactile Sensor. *IEEE Sensors Journal*, 16(22), 7853-7863. doi: 10.1109/JSEN.2016.2605134.

Rana, A. & Duchaine, V. (2013, May). Improved Soft Dielectric for Highly Sensitive Capacitive Tactile Sensor. *In 2013 IEEE International Conference on Robotics and Automation : Workshop on Research Frontiers in Electronic Skin Technology (Karlsruhe, Germany*.

Rasouli, M., Chen, Y., Basu, A., Kukreja, S. L. & Thakor, N. V. (2018). An Extreme Learning Machine-Based Neuromorphic Tactile Sensing System for Texture Recognition. *IEEE Transactions on Biomedical Circuits and Systems*, 12(2), 313-325. doi: 10.1109/TBCAS.2018.2805721.

Ries, E. (2017). *The Startup Way: How Entrepreneurial Management Transforms Culture and Drives Growth*. Penguin Books Limited. Consulted at https://books.google.ca/books?id=PNCWDgAAQBAJ.

RightHand Robotics, R. R. (2016). RightHand Robotics. [Online; accessed 9-April-2016], Consulted at http://www.righthandrobotics.com/.

Rispal, S. (2014). *Discrimination de textures et quantification de rugosité par algorithme d'apprentissage*. (Master's thesis, École de technologie supérieure, Montréal).

Roberge, E. & Duchaine, V. (2017, Sep.). Detecting insertion tasks using convolutional neural networks during robot teaching-by-demonstration. *2017 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pp. 3210-3216. doi: 10.1109/IROS.2017.8206154.

Roberge, J. P., Rispal, S., Wong, T. & Duchaine, V. (2016, May). Unsupervised feature learning for classifying dynamic tactile events using sparse coding. *2016 IEEE International Conference on Robotics and Automation (ICRA)*, pp. 2675-2681. doi: 10.1109/ICRA.2016.7487428.

Robotics, R. (2020, January, 24). RightHand Labs - Robotic Grippers - Tactile Sensors. Consulted at https://www.labs.righthandrobotics.com/.

Romano, J., Hsiao, K., Niemeyer, G., Chitta, S. & Kuchenbecker, K. (2011). Human-Inspired Robotic Grasp Control With Tactile Sensing. *Robotics, IEEE Transactions on*, 27(6), 1067-1079. doi: 10.1109/TRO.2011.2162271.

Roska, B. & Werblin, F. (2001). Vertical interactions across ten parallel, stacked representations in the mammalian retina. *Nature*, 1, 583–57.

Roweis, S. T. & Saul, L. K. (2000). Nonlinear Dimensionality Reduction by Locally Linear Embedding. *Science*, 290(5500), 2323–2326. doi: 10.1126/science.290.5500.2323.

Sato, A., Funabora, Y., Doki, S. & Doki, K. (2017, Dec). Contact force distribution predictive control system for wearable robot with tactile sensors. *2017 11th Asian Control Conference (ASCC)*, pp. 1373-1378. doi: 10.1109/ASCC.2017.8287372.

Schneider, A., Sturm, J., Stachniss, C., Reisert, M., Burkhardt, H. & Burgard, W. (2009, Oct). Object identification with tactile sensors using bag-of-features. *2009 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pp. 243-248. doi: 10.1109/IROS.2009.5354648.

Shalizi, C. (2009). Nonlinear Dimensionality Reduction I: Local Linear Embedding. [Online; accessed 10-April-2016], Consulted at https://www.stat.cmu.edu/~cshalizi/350/lectures/14/lecture-14.pdf.

Shea, H., Shintake, J. & Floreano, D. (2016). Soft compliant gripper for safe manipulation of extremely fragile objects. *SPIE Newsroom*, 1–3. doi: 10.1117/2.1201603.006409.

Shinoda, H., Sasaki, S. & Nakamura, K. (2000). Instantaneous evaluation of friction based on ARTC tactile sensor. *Proceedings 2000 ICRA. Millennium Conference. IEEE International Conference on Robotics and Automation. Symposia Proceedings (Cat. No.00CH37065)*, 3, 2173-2178 vol.3. doi: 10.1109/ROBOT.2000.846350.

Shirafuji, S. & Hosoda, K. (2011, June). Detection and prevention of slip using sensors with different properties embedded in elastic artificial skin on the basis of previous experience. *Advanced Robotics (ICAR), 2011 15th International Conference on*, pp. 459-464. doi: 10.1109/ICAR.2011.6088598.

Simpkins, A. (2013). Robotic Tactile Sensing: Technologies and System (Dahiya, R.S. and Valle, M.; 2013) [On the Shelf]. *IEEE Robotics Automation Magazine*, 20(2), 107-107. doi: 10.1109/MRA.2013.2255515.

Skedung, L., Arvidsson, M., Chung, J., Stafford, C., Berglund, B. & Rutland, M. (2013). Feeling Small: Exploring the Tactile Perception Limits. *Scientific reports*, 3, 2617. doi: 10.1038/srep02617.

Smith, L. (2002, February, 26). A tutorial on Principal Components Analysis. Consulted at http://www.cs.otago.ac.nz/cosc453/student_tutorials/principal_components.pdf.

Song, S., Drotlef, D.-M., Majidi, C. & Sitti, M. (2017). Controllable load sharing for soft adhesive interfaces on three-dimensional surfaces. *Proceedings of the National Academy of Sciences*, 114(22), E4344–E4353. doi: 10.1073/pnas.1620344114.

Stassi, S., Valentina, C., Canavese, G. & Pirri, C. (2014). Flexible Tactile Sensing Based on Piezoresistive Composites: A Review. *Sensors (Basel, Switzerland)*, 14, 5296-332. doi: 10.3390/s140305296.

Stevens, J. C. (1982). Temperature can sharpen tactile acuity. *Perception and Psychophysics*, 31(6), 577 - 580.

Su, Z., Hausman, K., Chebotar, Y., Molchanov, A., Loeb, G. E., Sukhatme, G. S. & Schaal, S. (2015, Nov). Force estimation and slip detection/classification for grip control using a biomimetic tactile sensor. *2015 IEEE-RAS 15th International Conference on Humanoid Robots (Humanoids)*, pp. 297-303. doi: 10.1109/HUMANOIDS.2015.7363558.

TakkTile. (2016). Home - TakkTile. [Online; accessed 9-April-2016], Consulted at https://www.labs.righthandrobotics.com/takktile-sensors.

Tandon, A., Shukla, P. & Patel, H. (2015, 12). Review of Transduction Techniques for Tactile Sensors and a Comparative Analysis of Commercial Sensors. pp. 133–138.

Teng, X., Wu, B., Yu, W. & Liu, C. (2005). A hand gesture recognition system based on local linear embedding. *Journal of Visual Languages and Computing*, 16(5), 442 - 454.

doi: https://doi.org/10.1016/j.jvlc.2005.04.003. Special issue section on Context and Emotion Aware Visual Interaction - Part I, pages 383- 441.

Tenzer, Y., Jentoft, L. P. & Howe, R. D. (2014). The Feel of MEMS Barometers: Inexpensive and Easily Customized Tactile Array Sensors. *IEEE Robotics Automation Magazine*, 21(3), 89-95. doi: 10.1109/MRA.2014.2310152.

Tincani, V., Catalano, M. G., Farnioli, E., Garabini, M., Grioli, G., Fantoni, G. & Bicchi, A. (2012, Oct). Velvet fingers: A dexterous gripper with active surfaces. *2012 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pp. 1257-1263. doi: 10.1109/IROS.2012.6385939.

Tincani, V., Grioli, G., Catalano, M. G., Garabini, M., Grechi, S., Fantoni, G. & Bicchi, A. (2013, May). Implementation and control of the Velvet Fingers: A dexterous gripper with active surfaces. *2013 IEEE International Conference on Robotics and Automation*, pp. 2744-2750. doi: 10.1109/ICRA.2013.6630955.

Tiwana, M. I., Redmond, S. J. & Lovell, N. H. (2012). A review of tactile sensing technologies with applications in biomedical engineering. *Sensors and Actuators A: Physical*, 179, 17 - 31. doi: https://doi.org/10.1016/j.sna.2012.02.051.

Tomo, T. P., Somlor, S., Schmitz, A., Jamone, L., Huang, W., Kristanto, H. & Sugano, S. (2016). Design and Characterization of a Three-Axis Hall Effect-Based Soft Skin Sensor. *Sensors*, 16(4), 1–11. Consulted at https://www.mdpi.com/1424-8220/16/4/491.

Tremblay, M. R. & Cutkosky, M. R. (1993, May). Estimating friction using incipient slip sensing during a manipulation task. *Robotics and Automation, 1993. Proceedings., 1993 IEEE International Conference on*, pp. 429-434 vol.1. doi: 10.1109/ROBOT.1993.292018.

University of California San Diego,Carnegie Mellon University, C. U. C. U. G. I. o. T. N. U. N. U. O. S. U. S. I. T. A. U. T. U. o. U. U. o. C. B. U. o. N. R. U. o. S. C. U. o. T. K. U. o. W. U. o. W. V. U. Y. U. (2016). *A Roadmap for U.S. Robotics: From Internet to Robotics*. Consulted at https://cra.org/ccc/wp-content/uploads/sites/2/2016/11/roadmap3-final-rs-1.pdf.

van der Maaten, L. & Hinton, G. (2008). Visualizing Data using t-SNE. *Journal of Machine Learning Research*, 9, 2579–2605. Consulted at http://www.jmlr.org/papers/v9/vandermaaten08a.html.

Wan, Q., Adams, R. P. & Howe, R. D. (2016, May). Variability and predictability in tactile sensing during grasping. *2016 IEEE International Conference on Robotics and Automation (ICRA)*, pp. 158-164. doi: 10.1109/ICRA.2016.7487129.

Ward-Cherrier, B., Pestell, N., Cramphorn, L., Giannaccini, M., Winstone, B., Rossiter, J. & Lepora, N. (2018). The TacTip Family: Soft Optical Tactile Sensors with 3D-Printed Biomimetic Morphologies. *Soft Robotics*, 5(2), 216–227. doi: 10.1089/soro.2017.0052.

Weerakkody, N. S., Mahns, D. A., Taylor, J. L. & Gandevia, S. C. (2007). Impairment of human proprioception by high-frequency cutaneous vibration. *The Journal of Physiology*, 581(3), 971–980.

Westling, G. & Johansson, R. S. (1984). Factors influencing the force control during precision grip. *Experimental Brain Research*, 53(2), 277–284. doi: 10.1007/BF00238156.

WRS, W. R. S. (2018, October, 17). World Robot Summit: Assembly Challenge - Industrial Robotics Category. Consulted at https://worldrobotsummit.org/en/wrc2018/industrial/assembly.html.

Wu, X. A., Suresh, S. A., Jiang, H., Ulmen, J. V., Hawkes, E. W., Christensen, D. L. & Cutkosky, M. R. (2015). Tactile sensing for gecko-inspired adhesion. *Intelligent Robots and Systems (IROS), 2015 IEEE/RSJ International Conference on*, pp. 1501–1507.

Wyk, K. V. & Falco, J. (2018). Slip Detection: Analysis and Calibration of Univariate Tactile Signals. *CoRR*, abs/1806.10451, 1-8.

Xie, H., Liu, H., Luo, S., Seneviratne, L. D. & Althoefer, K. (2013, Nov). Fiber optics tactile array probe for tissue palpation during minimally invasive surgery. *2013 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pp. 2539-2544. doi: 10.1109/IROS.2013.6696714.

Yang, Y.-J., Cheng, M.-Y., Chang, W.-Y., Tsao, L.-C., Yang, S.-A., Shih, W.-P., Chang, F.-Y., Chang, S.-H. & Fan, K.-C. (2008). An integrated flexible temperature and tactile sensing array using PI-copper films. *Sensors and Actuators A: Physical*, 143(1), 143 - 153. doi: https://doi.org/10.1016/j.sna.2007.10.077. Micromechanics Section of Sensors and Actuators (SAMM), based on contributions revised from the Technical Digest of the IEEE 20th International Conference on Micro Electro Mechanical Systems (MEMS 2007).

Yussof, H., Wada, J. & Ohka, M. (2009, May). Object handling tasks based on active tactile and slippage sensations in a multi-fingered humanoid robot arm. *Robotics and Automation, 2009. ICRA '09. IEEE International Conference on*, pp. 502-507. doi: 10.1109/ROBOT.2009.5152367.

Zhang, Y., Kan, Z., Tse, Y. A., Yang, Y. & Wang, M. Y. (2018). FingerVision Tactile Sensor Design and Slip Detection Using Convolutional LSTM Network. *CoRR*, abs/1810.02653, 1-7.

Zinn, M., Khatib, O., Roth, B. & Salisbury, J. K. (2004). Playing it safe [human-friendly robots]. *IEEE Robotics & Automation Magazine*, 11(2), 12–21.