DenseTact: Robots with a Human Touch
About Me

Internships

Cuyahoga Community College
Illinois
MITRE
Meyerhoff (M20)
LACO

Highschool
2008

Undergrad:
Mechanical Engineering

2012

Graduate:
MS – Robotics; PhD - MEAM

2019

Assistive Robotics and Manipulation Laboratory

Stanford University
Mission

The mission of the Assistive Robotics and Manipulation Lab is to develop intelligent, assistive, collaborative robots that improve human life.
Assistive Robotics and Manipulation Laboratory (ARMLab)
Outline

• Robotic Perception
• What is in a Touch?
• DenseTact
• The Future with Robotic Touch
Robotic Perception
Robotics Fundamentals

Intelligence/Planning

Controls/Actuation

Think → Act

See

Perception/Observation

\[ y_n = \max(0, W_{10 \times 2}x + b_{2 \times 1}) \]
\[ y_{n-1} = \max(0, W_{10 \times 10}y_n + b_{10 \times 1}) \]
\[ x = W_{2 \times 10}y_{n-1} + b_{2 \times 1} \]
Usefulness of Humanoid Robots

A few key markets:

Assisted Living for Older Persons

Rapid Manufacturing

Robohub.org

The Hindu BusinessLine
Why are humanoid robots not in homes yet?

Where we are

Honda 2006

DARPA Robotics Challenge 2015

Boston Dynamics 2021

Where we want to be

Star Trek: Data (AI Lifeform)

Over the past few decades, we have seen advances in robotic locomotion, vision, kinematics and control.

One of the key outstanding challenges is robotic manipulation - the ability to obtain and understand tactile feedback, and use it for planning and control the way humans can when performing the same tasks.
Future of Collaborative Robotics

The importance of Touch

Detecting incipient slip

Determining the pose of small objects within the hand and applied force

Manipulating soft and deformable objects

In-hand manipulation

https://www.easylatchunky.com/what-is-a-pcb-socket.htm
What is in a touch?
What types of grasps are there?

What does it mean to **grasp successfully**?

- Object is immobilized (robust to external disturbances)
- The object can be repositioned based on the high-level task

Rigid versus Soft Contact

Compliance is relative between the object and finger
Robotic Grasping in Industry

Robotic Grasping in Industry

Fanuc Robot: https://youtu.be/HoT7s7ujN6E
Assessing Grasp Quality

Form Closure

Object is geometrically incapable of moving
*(Shape sensing)*

Force Closure

Object is immobilized by both finger presence, as well as friction forces capable of rejecting external forces (e.g. gravity)
*(Shape and Force sensing)*

Not being able to assess grasp quality makes it harder to predict and recover from possible failure modes!
What forces could be sensed?

If you had the perfect sensor, which forces could be sensed?

Sensor Surface

$\sigma_{33} = \sigma_z$

Normal force

$\sigma_{12} = \sigma_{21} = \tau_{xy}$

Planer shear forces

$\sigma_{11} = \sigma_x$

Planer torsion force

$\sigma_{22} = \sigma_y$
**Touch in Humans**

**Tactile Receptors in the Skin**

Human biology leverages physical transduction (strain on cells is converted to electrical impulses and sent to the brain).

<table>
<thead>
<tr>
<th>Mechanoreceptors</th>
<th>Type (depth)</th>
<th>Type (adapting speed)</th>
<th>Receptive Field</th>
<th>Frequency (Hz)</th>
<th>Sensing Property</th>
</tr>
</thead>
<tbody>
<tr>
<td>Merkel's discs</td>
<td>I</td>
<td>Slow</td>
<td>Small (3-4mm)</td>
<td>5-15</td>
<td>Pressure, Texture</td>
</tr>
<tr>
<td>Ruffini endings</td>
<td>II</td>
<td>Slow</td>
<td>Large (&gt;10mm)</td>
<td>15-400</td>
<td>Stretch</td>
</tr>
<tr>
<td>Meissner's corpuscles</td>
<td>I</td>
<td>Fast</td>
<td>Small (3-4mm)</td>
<td>20-50</td>
<td>Stroke, Fluttering</td>
</tr>
<tr>
<td>Pacinian corpuscles</td>
<td>II</td>
<td>Fast</td>
<td>Large (&gt;20mm)</td>
<td>60-400</td>
<td>Vibration</td>
</tr>
</tbody>
</table>

Touch in Humans (coupled with vision?)

Tactile information travels to the Somatosensory Cortex. Area 3b is responsible for processing basics of touch sensations.

Kuehn et al. showed in 2018 that the fine-grained finger maps in human primary somatosensory cortex area 3b are activated not only during tactile mechanical stimulation, but also when viewing the same fingers being touched. Hence, viewing (hypothetical) touches activated the same region of the brain as the true tactile sensation.

Our brains recognize the advantage in correlated vision with tactile sensation, but biology has not evolved to let us fully exploit that relationship!

Types of Robotic Tactile Sensors

Physical Transduction

Material is deformed and as a result of that deformation an electrical signal is sent to a processor and is correlated to force or shape change.

**Challenges:**
- a) obtaining multiple modalities
- b) cross-talk when embedded in soft medium

Optical-based

Soft material is deformed, and a camera (optics) observes the deformation and correlates change in boundary deformation to shape and forces.

**Challenges:**
- a) field of view of the camera may be limited
- b) characterizing the material and expected image for strain/stress at high resolution
Types of Robotic Tactile Sensors

Physical Transduction

Single capacitive sensor
(normal force detected over entire plate)

Array of nibs that leverage relative capacitance to detect normal, shear, torsion through groups of contacts.

Fig. 2. Cross sectional and isometric view of the proposed design. (a) Idle state. (b) Under normal compression. (c) Under shear force. (d) Vibration under slippage. (e) Overall design.

Types of Robotic Tactile Sensors

Optical-based Tactile Sensors

2D-shaped vision-based sensor

DIGIT, Gelsight
Limited grasping manipulation because of 2D-gel surface

3D-shaped vision-based sensor

Omnitact
Design cost ($3200) because of micro-cameras

The Holy Grail for Robotic Touch

The gap in existing sensor need to address the following at once [1]:

1) High resolution
2) Highly sensitive
3) Reliable
4) Easy to use
5) Compact
6) Inexpensive


DenseTact: Calibrated Optical Tactile Robotic Touch

Do, Won Kyung, and Monroe Kennedy III. "DenseTact: Optical Tactile Sensor for Dense Shape Reconstruction." IEEE ICRA 2022 (Accepted)
DenseTact

Motivation

Improving robotic dexterous manipulation is the catalyst to enabling ubiquitous robots capable of performing advanced collaborative tasks.

Robots capable of manipulating small objects with the ability accurately assess stability and adapt or re-grasp is necessary for systems to be effective in tasks ranging from rapid industrial assembly to assisted living tasks.

DenseTact is an optical tactile sensor whose first generation provides calibrated, high-resolution shape reconstruction.
DenseTact: Optical Tactile Sensor for Dense Shape Reconstruction

Won Kyung Do¹, Monroe Kennedy III¹

¹Stanford University, Stanford, CA, USA
DenseTact Design

Related Work

Fig. 3. Exploded diagram of the GelSlim 3.0 tactile finger and its 10 components


Fig. 2. Exploded view of a single DIGIT sensor. A) elastomer, B) acrylic window, C) snap-fit holder, D) lighting PCB, E) plastic housing, F) camera PCB, G) back housing.


Design commonalities

- Elastomer gel (often with coating, possibly markers on boundary)
- LED illuminated interior housing
- Camera for sensor surface observation
- Sensor Housing and off-site computation

Elastomer gel, LED illuminated interior housing, Camera for sensor surface observation, and Sensor Housing and off-site computation are design commonalities of various tactile sensors.

DenseTact Design

Related Work

<table>
<thead>
<tr>
<th>Name</th>
<th>3D shape</th>
<th>Resolution</th>
<th>Calibration</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gelsight-like sensors [20]–[22]</td>
<td>×</td>
<td>640 × 480</td>
<td>3D</td>
</tr>
<tr>
<td>Softbubble [13]</td>
<td>✓</td>
<td>224 × 171</td>
<td>3D</td>
</tr>
<tr>
<td>Omnitact [23]</td>
<td>✓</td>
<td>400 × 400</td>
<td>×</td>
</tr>
<tr>
<td>NeuTouch [24]</td>
<td>×</td>
<td>39</td>
<td>×</td>
</tr>
<tr>
<td>TacTip [25]</td>
<td>✓</td>
<td>127 180</td>
<td>×</td>
</tr>
<tr>
<td>Optofiber-sensor [26]</td>
<td>×</td>
<td>61 fibers</td>
<td>2D</td>
</tr>
<tr>
<td>FingerVision [27]</td>
<td>×</td>
<td>640 × 480</td>
<td>3D</td>
</tr>
<tr>
<td>GelTip [28]</td>
<td>✓</td>
<td>not specified</td>
<td>2D</td>
</tr>
<tr>
<td><strong>This work: DenseTact</strong></td>
<td>✓</td>
<td>800 × 600</td>
<td>3D</td>
</tr>
</tbody>
</table>

TABLE I: Related Work. List of high-resolution vision-based tactile sensors for shape reconstruction. 2D-calibrated sensors localize the surface position without estimating depth.

DenseTact Design

**Elastomer**: Extra-soft silicone (P-565 Platinum Clear Silicone 20:1 ratio. It has a 6.5 Shor-A hardness (similar to human skin).

**Reflective gel**: *Inhibit X* is applied as an adhesive before airbrushing the surface with mixture of reflective metallic ink and silicone (*Smooth-on Psycho Paint*).

**Camera and Illumination**: The camera is a Sony IMX179 image sensor (8MP, 30fps) with a 185 deg field of view. The interior is illuminated with LED strip (3 are used at a time).

**Chassis and Final Cost**: The chassis is 3D printed (PLA), and the cost for the entire sensor is $80 (camera system is $70, LEDs $5.5, Elastomer $3.5, Camera mount $1).
DenseTact
Modeling and Calibration

Camera Calibration → Shape Reconstruction

- Modeling Principle
- Calibration Technique
DenseTact
Visualization - FishEye Lens

Early versions of DenseTact (2020) used multiple cameras to visualize the interior; this proved to be large and computationally expensive.

We realized the value of using fish-eye lenses as they would allow for full visualization of the sensor interior, the hemispherical shape.
Further, we could leverage the equiangular projection of a fisheye lens within our hemispherical sensor, allowing for an (almost) undistorted view.

Superimposed fisheye projection with orthographic rendering from an orthographic mirror.

Picture Credit: Paul Bourke (http://paulbourke.net/dome/fisheye/)
To calibrate the fisheye lens for this task, we needed to account for the distortion caused by the fisheye lens even with equiangular projection.

For variables $R$ being the radius of the hemisphere, $\theta$ being the pitch, and $\psi$ being the azimuth, the distortion is symmetric with respect to $\psi$, but varies with respect to the projection of $R \sin(\theta)$.

This is modeled with a Gaussian Process (GP) to provide the inverse function:

$$ R \sin(\theta) = f_{GP}(r(u, v)) $$

$$(\theta, \phi) = \left( \sin^{-1} \left( \frac{f_{GP}(r)}{R} \right), \tan^{-1} \left( \frac{v - v_c}{u - u_c} \right) \right)$$
DenseTact
Shape Reconstruction

\[ N(x, y) = \left( \frac{\partial f}{\partial x}, \frac{\partial f}{\partial y}, -1 \right) \]

\[ z = f(x, y) \]

Light intensity from internal reflectance directly correlates the surface normal

\[ I(u, v) = L \left( \frac{\partial f}{\partial u}(u, v), \frac{\partial f}{\partial v}(u, v), u, v \right) \]

So the goal is to learn an inverse function that maps intensity to the surface normal

\[ \left( \frac{\partial f}{\partial x}, \frac{\partial f}{\partial y} \right) = L^{-1}(I(u, v)) \]

Then leveraging Poisson integration (with known boundary conditions) allows us to solve for the surface shape:

\[ \nabla^2 f = g = \frac{\partial}{\partial x} \left( \frac{\partial f}{\partial x} \right) + \frac{\partial}{\partial y} \left( \frac{\partial f}{\partial y} \right) \]

Most other work accomplishes this through look up tables, we propose to solve this with a network

\[ (R, \theta, \psi) = M(I_{rgb}(u, v)) \]
DenseTact
Shape Reconstruction

To map the entire possible contact area at once, we push known shapes (3D printed parts with 0.1mm accuracy) into the sensor.

We used 37 different hemispherical indentors, a total of 25 different sub-indentors with various shapes were also printed and combined in different configurations with the hemispherical indentors (changing angles A,B,C).

Leveraging ray tracing, we can produce the expected radial depth image to be produced.

\[ R_{ray}(u, v) = f_{raycast}(Mesh_{stl}, \theta(u, v), \psi(u, v)) \]
DenseTact
Shape Reconstruction - Modeling

Aspects of our model:

• We realize the importance of boundary conditions informing the entire depth image (Poisson)

• The CNN autoencoder structure helps ensure compactness of internal model, while skip-connections provide high-resolution

• 3D parts are used with recasting for supervised training

The sensor interior is the input to the autoencoder network. The ground truth is provided from the object CAD model and converted to a depth image. The resultant disparity between prediction and ground truth is used to train the network. The output depth image is converted to a 3D point-cloud via a correspondence step.
DenseTact

Results - Shape Reconstruction

Input Image

Ground Truth

Model Prediction
DenseTact
Results - Shape Reconstruction

The red line refers to the error of the ground truth, which is 109.6 micron from the precision error of the 3D printer. The model is justified by specifying the error of ground truth which matters while it goes to the 100-micron scale. The mean of $L_1$ loss for training and test set is 0.2381mm and 0.2811mm, respectively. The mean of $L_2$ loss for all training and test sets is 0.0306mm and 0.03208mm. In other words, the DenseTact sensor performs the shape reconstruction with an absolute mean error of 0.28mm.

Each data point represents the mean $L_1$ re-projection error for a single image (effectively 253,213 pixels). Statistics are shown for the training (29,200 images) and test sets (1000 images).
DenseTact

Results - In-hand Localization

Iterative Closest Point (ICP) is used to estimate the pose of the object held in the hand. We evaluated the sensor by measuring first RMS error and fitness between detected DenseTact point cloud and corresponding points from object point cloud. Fitness refers to the ratio of the number of inlier correspondences and the number of DenseTact point cloud.

After 23 grasping trial, the average fitness score was 0.597 ($\sigma = 0.238$) and average RMS error was 0.037184 ($\sigma = 0.00276$). Note that the average RMS error after 200 iterations of ICP was 0.0211.
DenseTact
2nd Generation

$F_S$

Marker Displacement

Normal
Shear
Torsion
DenseTact
2nd Generation
DenseTact
2nd Generation
DenseTact
2nd Generation

Physics Simulator

Gaussian Process (GP)
Model for Upsampling

\[
[F^* \quad \sigma^*] = GP_\theta(x^*, x, F)
\]

\(\theta\): kernel parameters

Fingertip RGB Image (size \((m \times n \times 3)\))

- External calibration sensor (array)
- GP model certainty

Stanford University
DenseTact
2nd Generation

Detecting forces allows us to characterize the wrench cone limit surface as well as the position within the limit surface.
The Future with Robotic Touch
Touch to Explore
Visual Tactile Neural Fields

Measurement (Task I.B)

Physical Object Library

Action Selection (Thrust II)

Visual Tactile Neural Field (Task I.A)

https://youtu.be/JuH79E8rdKc
Touch to Explore
Visual Tactile Neural Fields

\[ r(p) = C_i + td \]

\[ r, g, b; color \]

\[ C_i \]

\[ \rho; radius \]

\[ p \rightarrow MLP_{\text{position}} \rightarrow \rho \]

\[ e \rightarrow MLP_{\text{rgb}} \rightarrow c \]

Denseact Image

Object Properties

- MLP → Color
- MLP → Surface normal
- MLP → Texture
In-Hand Manipulation
Finger Gait Control with Knowledge of Stability

Detecting incipient slip

In-hand manipulation
Learn to Touch
LfD for Force Control Tasks

Determining the pose of small objects within the hand and applied force

https://www.easytechjunkie.com/what-is-a-pcb-socket.htm
Soft Object Manipulation

Manipulating soft and deformable objects
Thank you!