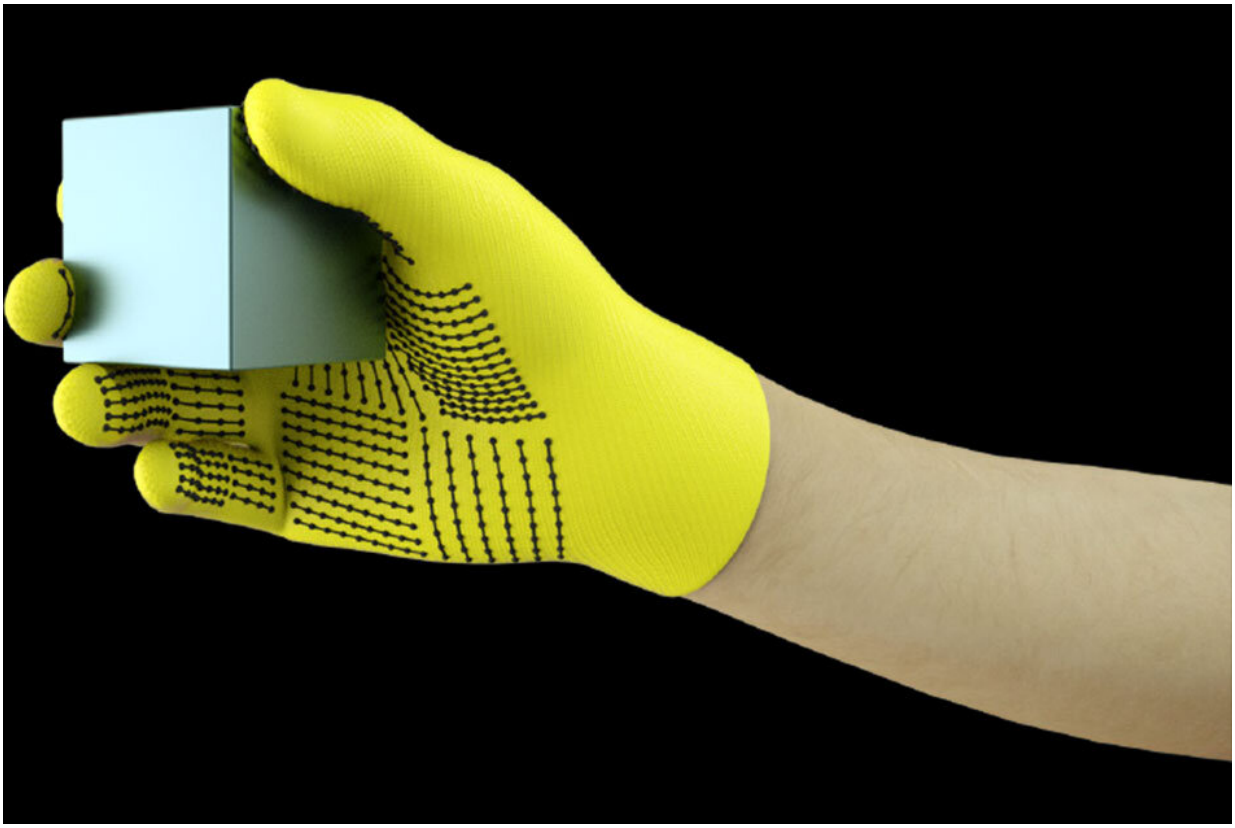
Sensor-packed glove learns signatures of the human grasp

May 29 2019, by Rob Matheson



MIT researchers have developed a low-cost, sensor-packed glove that captures pressure signals as humans interact with objects. The glove can be used to create high-resolution tactile datasets that robots can leverage to better identify, weigh, and manipulate objects. Credit: Massachusetts Institute of Technology

Wearing a sensor-packed glove while handling a variety of objects, MIT

researchers have compiled a massive dataset that enables an AI system to recognize objects through touch alone. The information could be leveraged to help robots identify and manipulate objects, and may aid in prosthetics design.

The researchers developed a low-cost knitted glove, called "scalable tactile glove" (STAG), equipped with about 550 tiny sensors across nearly the entire hand. Each sensor captures pressure signals as humans interact with objects in various ways. A neural network processes the signals to "learn" a dataset of pressure-signal patterns related to specific objects. Then, the system uses that dataset to classify the objects and predict their weights by feel alone, with no visual input needed.

In a paper published in *Nature*, the researchers describe a dataset they compiled using STAG for 26 common objects—including a soda can, scissors, tennis ball, spoon, pen, and mug. Using the dataset, the system predicted the objects' identities with up to 76 percent accuracy. The system can also predict the correct weights of most objects within about 60 grams.

Similar sensor-based gloves used today run thousands of dollars and often contain only around 50 sensors that capture less information. Even though STAG produces very high-resolution data, it's made from commercially available materials totaling around $10.

The tactile sensing system could be used in combination with traditional computer vision and image-based datasets to give robots a more human-like understanding of interacting with objects.

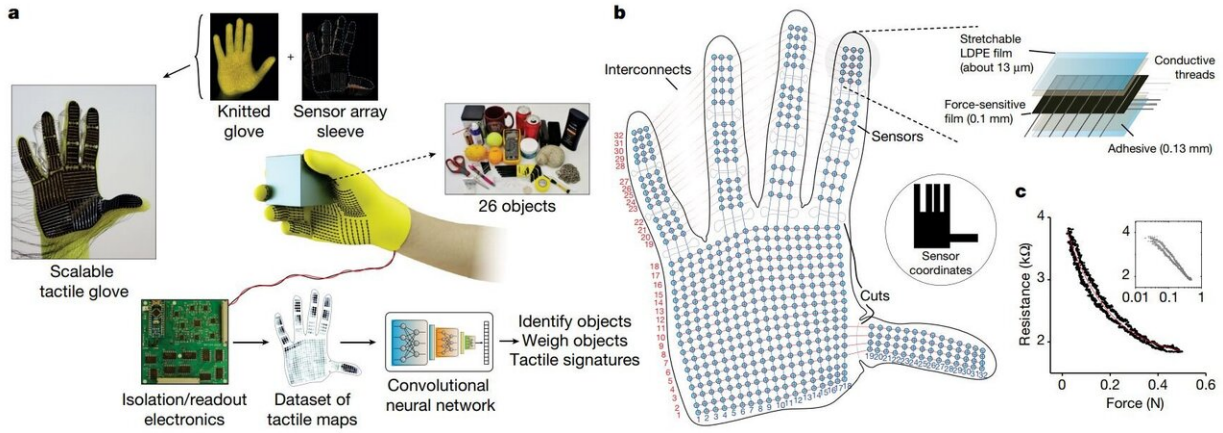
"Humans can identify and handle objects well because we have tactile feedback. As we touch objects, we feel around and realize what they are. Robots don't have that rich feedback," says Subramanian Sundaram Ph.D. '18, a former graduate student in the Computer Science and

Artificial Intelligence Laboratory (CSAIL). "We've always wanted robots to do what humans can do, like doing the dishes or other chores. If you want robots to do these things, they must be able to manipulate objects really well."

The researchers also used the dataset to measure the cooperation between regions of the hand during object interactions. For example, when someone uses the middle joint of their index finger, they rarely use their thumb. But the tips of the index and middle fingers always correspond to thumb usage. "We quantifiably show, for the first time, that, if I'm using one part of my hand, how likely I am to use another part of my hand," he says.

Prosthetics manufacturers can potentially use information to, say, choose optimal spots for placing pressure sensors and help customize prosthetics to the tasks and objects people regularly interact with.

Joining Sundaram on the paper are: CSAIL postdocs Petr Kellnhofer and Jun-Yan Zhu; CSAIL graduate student Yunzhu Li; Antonio Torralba, a professor in EECS and director of the MIT-IBM Watson AI Lab; and Wojciech Matusik, an associate professor in electrical engineering and computer science and head of the Computational Fabrication group.



The STAG as a platform to learn from the human grasp. Credit: *Nature* (2019).
DOI: 10.1038/s41586-019-1234-z

STAG is laminated with an electrically conductive polymer that changes resistance to applied pressure. The researchers sewed conductive threads through holes in the conductive polymer film, from fingertips to the base of the palm. The threads overlap in a way that turns them into pressure sensors. When someone wearing the glove feels, lifts, holds, and drops an object, the sensors record the pressure at each point.

The threads connect from the glove to an external circuit that translates the pressure data into "tactile maps," which are essentially brief videos of dots growing and shrinking across a graphic of a hand. The dots represent the location of pressure points, and their size represents the force—the bigger the dot, the greater the pressure.

From those maps, the researchers compiled a dataset of about 135,000 video frames from interactions with 26 objects. Those frames can be used by a neural network to predict the identity and weight of objects, and provide insights about the human grasp.

To identify objects, the researchers designed a convolutional neural network (CNN), which is usually used to classify images, to associate specific pressure patterns with specific objects. But the trick was choosing frames from different types of grasps to get a full picture of the object.

The idea was to mimic the way humans can hold an object in a few different ways in order to recognize it, without using their eyesight.

Similarly, the researchers' CNN chooses up to eight semirandom frames from the video that represent the most dissimilar grasps—say, holding a mug from the bottom, top, and handle.

But the CNN can't just choose random frames from the thousands in each video, or it probably won't choose distinct grips. Instead, it groups similar frames together, resulting in distinct clusters corresponding to unique grasps. Then, it pulls one frame from each of those clusters, ensuring it has a representative sample. Then the CNN uses the contact patterns it learned in training to predict an object classification from the chosen frames.

"We want to maximize the variation between the frames to give the best possible input to our network," Kellnhofer says. "All frames inside a single cluster should have a similar signature that represent the similar ways of grasping the object. Sampling from multiple clusters simulates a human interactively trying to find different grasps while exploring an object."

For weight estimation, the researchers built a separate dataset of around 11,600 frames from tactile maps of objects being picked up by finger and thumb, held, and dropped. Notably, the CNN wasn't trained on any frames it was tested on, meaning it couldn't learn to just associate weight with an object. In testing, a single frame was inputted into the CNN. Essentially, the CNN picks out the pressure around the hand caused by the object's weight, and ignores pressure caused by other factors, such as hand positioning to prevent the object from slipping. Then it calculates the weight based on the appropriate pressures.

The system could be combined with the sensors already on robot joints that measure torque and force to help them better predict object weight. "Joints are important for predicting weight, but there are also important components of weight from fingertips and the palm that we capture,"

Sundaram says.

More information: Learning the signatures of the human grasp using a scalable tactile glove, *Nature* (2019). DOI: 10.1038/s41586-019-1234-z , www.nature.com/articles/s41586-019-1234-z

This story is republished courtesy of MIT News (web.mit.edu/newsoffice/), a popular site that covers news about MIT research, innovation and teaching.

Provided by Massachusetts Institute of Technology