Tactile Intelligence: Why it’s the Future of Robotic Grasping

The challenge: creating robots that can grasp any object

The simple task of picking something up is not as easy as it seems. Not for a robot, at least. Roboticists aim to develop a robot that can pick up anything – but today most robots perform “blind grasping,” where they’re dedicated to picking up an object from the same location every time. If anything changes, such as the shape, texture, or location of the object, the robot won’t know how to respond, and the grasp attempt will most likely fail.

Robots are still a long way off from being able to grasp any object perfectly on their first attempt. Why do grasping tasks pose such a difficult problem? Well, when people try to grasp something they use a combination of senses, the primary ones being visual and tactile. But so far, most attempts at solving the grasping problem have focused on using vision alone.

This approach is unlikely to give results that fully match human capabilities, because although vision is important for grasping tasks (such as for aiming at the right object), vision simply cannot tell you everything you need to know about grasping. Consider how Steven Pinker describes all the things the human sense of touch accomplishes: “Think of lifting a milk carton. Too loose a grasp, and you drop it; too tight, and you crush it; and with some gentle rocking, you can even use the tugging on your fingertips as a gauge of how much milk is inside!” (How the Mind Works, page 12). Because robots lack these sensing capabilities, they still lag far behind humans when it comes to even the simplest pick-and-place tasks.
In this post, you’ll learn more about why the current focus on robotic vision is unlikely to enable perfect grasping. As you’ll see from our exploration of recent developments in grasping methods, the future of robotic grasping requires tactile intelligence to work alongside vision.

**Previous studies have focused on vision, not tactile intelligence**

So far, most of the research in robotic grasping has aimed at building intelligence around visual feedback. One way of doing so is through database image matching, which is the method used in the Million Objects Challenge at Brown’s Humans to Robots Lab (http://h2r.cs.brown.edu/million-object-challenge/). The idea is for the robot to use a camera to detect the target object and monitor its own movements as it attempts to grasp the object. While doing so, the robot compares the real-time visual information with 3D image scans stored in the database. Once the robot finds a match, it can find the right algorithm for its current situation.

But while Brown’s approach aims to collect visual data for a variety of objects, roboticists are unlikely to ever develop a visual database for every single item a robot could possibly encounter. Moreover, the database-matching approach doesn’t include environmental constraints, so it doesn’t let the robot adapt its grasp strategy for different contexts.

Other researchers have turned to machine learning techniques for improving robotic grasping. These techniques allow robots to learn from experience, so eventually the robots can figure out the best way to grasp something on their own. Plus, unlike the database-matching methods, machine learning requires minimal prior knowledge. The robots don’t need to access a pre-made image database – they just need plenty of practice.

As Evan Ackerman reported for IEEE Spectrum (http://spectrum.ieee.org/automaton/robotics/artificial-intelligence/google-large-scale-robotic-grasping-project), Google recently conducted an experiment in grasping technology that combined a vision system with machine learning. In the past, researchers tried to improve grasping abilities by teaching the robots to follow whatever method the humans had decided was best. Google’s biggest breakthrough was in showing how robots could teach themselves – using a deep CNN, a vision system, and a lot of data (from 800,000 grasp attempts) – to improve based on what they learned from past experiences.
Their results seem extremely promising: since the robots’ responses were not pre-programmed, all of their improvements can truly be said to have “emerged naturally from learning” (as one of their researchers put it). But there are limits to what vision can tell robots about grasping, and Google might have already reached that frontier.

**Focusing solely on vision leads to certain problems**

There are three main reasons why the challenges Google and others are working on are hard to overcome with vision alone. First, vision is subject to numerous technical limitations. Even state-of-the-art vision systems can have trouble perceiving objects in certain light conditions (such as translucidity, reflection, and low-contrast colors), or when the object is too thin.

Second, many grasping tasks involve scenarios where it’s hard to see the entire object, so vision often cannot provide all the information the robot might need. If a robot is trying to pick up a wooden block from a table, a simple vision system would only observe the top of the block, and the robot would have no idea what the block looks like on the other side. In a more complex task like bin-picking, which involves multiple objects, the target object could be partially or fully obscured by surrounding items.

Finally, and most importantly, vision is simply not suited to the nature of the problem: grasping tasks are a matter of contact and forces, which cannot be monitored by vision. At best, vision can inform the robot about finger configurations that are likely to succeed, but in the end a robot needs tactile information to know the physical values that are associated with grasping tasks.

**How tactile intelligence can help**

The sense of touch plays a central role for humans during grasping and manipulation tasks. For amputees that have lost their hands, one of the biggest sources of frustration is the inability to sense what they’re touching while using prosthetic devices (Atkins et. al, “Prosthetic use in adult upper limb amputees: a comparison of the body powered and electrically powered prostheses”). Without the sense of touch, the amputees have to pay close visual attention during grasping and manipulation tasks, whereas a non-amputee could pick something up without even looking at it.

Researchers are aware of the crucial role that tactile sensors play in grasping, and the past 30 years have seen many attempts at building a tactile
sensor that replicates the human apparatus. However, the signals sent by a tactile sensor are complex and of high dimension, and adding sensors to a robotic hand often doesn’t directly translate into improved grasping capabilities. What’s needed is a way to transform these raw and low-level data into high-level information that will result in better grasping and manipulation performance. Tactile intelligence could then give robots the ability to predict grasp success using touch, recognize object slippage, and identify objects based on their tactile signatures.

At ÉTS’s CoRo Lab, researchers are building the blocks that will form the core of this new intelligence. One of the most recent developments is a machine learning algorithm that uses pressure images to predict successful and failed grasps. Developed by students Deen Cockburn and Jean-Philippe Roberge the system is an attempt to bring robots closer to human-level grasping abilities. We humans, of course, learn to recognize whether the configuration of our fingers will lead to a good grasp by using our sense of touch. Then we adjust our finger configuration until we’re confident our grasp attempt will succeed. Before robots can learn to quickly adjust their configurations, they need to get better at predicting the outcome of a grasp.

Photo 1: The tactile data were obtained during attempts to grasp various everyday objects. The objects were selected based on a list published by Amazon for the Amazon Picking Challenge 2015 (http://amazonpickingchallenge.org/item_list.txt).
This is where the CoRo Lab has triumphed. By combining a robotic hand from Robotiq, Inc. with a UR10 manipulator from Universal Robot, and adding several multimodal tactile sensors they built in-house and a Kinect-based vision system (used only to aim at the geometrical center of each object), the resulting robot was able to pick up a variety of objects and use this data to learn. Ultimately, the researchers succeeded at creating a system that correctly predicts grasp failure 83% of the time.

![Photo 2](image)

*Photo 2*: One of the robotic hands used for the experiments – a three-finger gripper from Robotiq, Inc. Another experiment used a two-finger gripper (not pictured) also from Robotiq.

Another team at the CoRo Lab, this time led by Jean-Philippe Roberge, focused specifically on slippage detection (Roberge et. al, “Unsupervised feature learning for classifying dynamic tactile events using sparse coding”). We humans can quickly recognize when an object is slipping out of our grasp because our fingers contain fast-adaptive mechanoreceptors, which are receptors in our skin that detect rapid changes in pressure and vibration. Object slippage generates vibrations on the surface of one’s hand, so the researchers input vibration images (spectrograms) instead of pressure images to
their machine learning algorithm. Using the same robot as the one from the grasp-prediction experiment, their system was able to learn the features of vibration images that correspond with object slippage, and it identified object slippage with 92% accuracy.

It may seem easy to get a robot to notice slippage, since slippage is just a series of vibrations. But how do you get the robot to learn the difference between vibrations that occur because the object is about to slip out of the robot’s grasp, and vibrations that occur because the robot is dragging the object across another surface (e.g. a table)? Don’t forget to take into account the tiny vibrations caused by the robot’s arm as it moves. These are three different events that generate similar signals, but require very different reactions from the robot. Distinguishing between these events is where machine learning comes in.

Photo 3: The tactile sensor built by the CoRo Lab is an ultrasensitive multimodal tactile sensor that can measure pressure, contact location, and vibration. Here, a water droplet rests on the surface of the tactile sensor – even its light weight generates a significant measurement.

When it comes to machine learning, the two CoRo teams have one thing in common: neither of them use handcrafted features for their machine learning algorithm. In other words, the system itself determines what is relevant for classifying slippage (or for predicting the grasp outcome, for the grasp-prediction experiment), instead of relying on the researchers’ guesses about what the best indicator might be.
“High-level features” in the past were always handcrafted, meaning the researchers would manually select the features that could help distinguish between types of slippage events (or between a good and bad grasp). For instance, they might decide that a pressure image showing the robot had grasped only the tip of the item would make for a bad grasp. But it’s actually far more accurate to let the robot learn on its own, since researchers’ guesses don’t always match reality.

Sparse coding is particularly useful for this purpose. It’s an unsupervised feature learning algorithm, and it works by creating a sparse dictionary that is used to represent new data. First, the dictionary creates itself by taking the spectrograms (or raw pressure images) as the input to a sparse coding algorithm. It outputs the dictionary, which consists of representations of high-level features. Then when new data arrives from subsequent grasp attempts, the dictionary is used to transform the new raw data into representations of that data, which are called sparse vectors. Finally, the sparse vectors are grouped into various causes of vibrations (or successful and failed grasp outcomes).

![Sparse dictionaries](image)

*Photo 4:* Sparse dictionaries (filtered) for the pressure images from the grasp-prediction experiment (left) and the spectrograms from the slippage detection experiment (right).

The researchers are currently testing ways to have the sparse coding algorithm automatically update itself, so that each grasp attempt will help the robot get better at making predictions. The idea is that eventually the robot will be able to use this information to adjust its behavior during the grasp. Ultimately, this research is a great example of how tactile and visual intelligence can work together to help robots learn how to grasp different objects.

**The future of tactile intelligence**

The key takeaway from this research is not that vision should be left by the wayside. Vision still makes crucial contributions to grasping tasks.
However, now that artificial vision has reached a certain level of development, it could be better to focus on developing new aspects of tactile intelligence, rather than continue to emphasize vision alone so strongly.

Jean-Philippe Roberge, researcher at the CoRo Lab, compares the potential of researching vision vs. tactile intelligence to Pareto’s 80-20 formula: now that the robotics community has the first 80% of visual intelligence down pat, perfecting the last 20% of vision is hard to do and won’t contribute much to object-manipulation tasks. By contrast, roboticists are still working on the first 80% of tactile sensing. So perfecting this first 80% will be relatively easy to do, and it has the potential to make a tremendous contribution to robots’ grasping abilities.

We may still be a long way off from the day when a robot can identify any object through touch, let alone clean your room for you – but when that day comes, we’ll surely have tactile intelligence researchers to thank.