Paper Review of GelSight

Alvin Sun

November 3, 2021

1 Paper Summary

This paper introduces a new way of high resolution tactile sensing via imaging techniques. The proposed GelSight sensors combines a single camera with a coated elastomer surface that can be illuminated with well controlled lighting conditions. Using multiple colors and lighting angles combined with the calibrated reflectance model, the captured image can be converted to surface normals at each pixel locations, and then ultimately integrates to a z-height map that fully recovers the geometry of the contacting portion of the object. In addition to sensing shapes, GelSight is also capable of estimating contact forces by exploiting the relationship with the motion field on the elastomer surface. The motion field is captured by tracking a dense grid of markers added to the elastomer. CNN based vision algorithms is then used to estimate the forces given marker motions. Slippage detection follows from a previous work which is by analyzing the statistics of the motion field. This work not only develops the theory, but also fabricated a family of GelSight sensors with different configurations. The experiments conducted with the actual sensors verify that this type of sensor can do what the paper claims with high fidelity.

2 What I Learned

1. Given some specific contact shape, the contact forces are almost linear in relation to the magnitude of the motion field.
2. In controlled lighting conditions, we can also do stereo photometric reconstruction with a single camera but multiple light sources.
3. The inverse of nonlinear projection models can usually be found via table lookup.

3 Opinions

3.1 Up Votes

I really like their stereo photometric method for finding the height map of the surface. It is really an elegant way to do 3D surface reconstruction with only monocular vision. Under controlled lighting and reflectance, when comparing to the more traditional multi-camera stereo vision, this could be not only easier and cheaper to fabricate but also more robust against noises as it does not require to solve the correspondence problem in stereo vision.

3.2 Down Votes

I don't quite agree with their method for force estimation. Even though CNN has been shown to be really high performing in image related tasks, predicting contact forces purely from motion fields images, in my opinion, is reliable enough for it to be deployed into unknown environment. As the authors mentioned, the shape of the object could influence the slope of the linear relationship between the force and the magnitude of the motion field. The contact motions are usually dynamic and time varying, so I believe modeling force estimation with time series of motion field changes as oppose to static motion field is more appropriate.

4 Evaluations

The goal of this paper is to develop a new type of tactile sensor that is not only capable of outputting high-resolution information about contact shape and forces, but also is straightforward to fabricate. This is a perfectly valid objective as very few previous work presents a tactile sensor design that captures this much of high-resolution details of each contact. Also, the fabrication details are usually neglected from previous work, which is usually nontrivial due to the design complexity. This paper certainly achieves their goal by presenting novel theories of vision based tactile sensing that also follows from detailed design with fabrication procedures.

The quality of this paper is really beyond exceptional in my opinion. It is really the first to make this rich tactile sensing data available for a wide range of applications. The only theoretical assumption they made is knowing the lighting conditions as well as the reflectance of the coating material. This assumption is relatively realistic given that this paper also provided calibration procedures for finding those prior knowledge. Not only the proposed method provides rich shape and contact force information, but also is quite feasible to fabricate into a relatively small-sized device that can be utilized in a robotic gripper setting. They also show quantitative evaluation results for real-world data, which further verifies the efficacy of their proposed methods.

5 Questions

1. If they used CNN for force / torque estimation, why did they not use it for slip detection?
2. I am wondering how this sensor is integrated with optimal planning of dexterous manipulation?