Bayesian Grasp: Vision based robotic stable grasp via prior tactile knowledge learning

Teng Xue, Wenhai Liu, Mingshuo Han, Zhenyu Pan, Jin Ma, Quanquan Shao, Weiming Wang*

***Abstract*— Robotic grasp detection is a fundamental capability for intelligent manipulation in unstructured environments. Previous work mainly employed visual and tactile fusion to achieve stable grasp, while, the whole process depending heavily on regrasping, which wastes much time to regulate and evaluate. We propose a novel way to improve robotic grasping: by using learned tactile knowledge, a robot can achieve a stable grasp from an image.**

First, we construct a prior tactile knowledge learning framework with novel grasp quality metric which is determined by measuring its resistance to external perturbations. Second, we propose a multi-phases Bayesian Grasp architecture to generate stable grasp configurations through a single RGB image based on prior tactile knowledge.

Results show that this framework can classify the outcome of grasps with an average accuracy of 86% on known objects and 70% on novel objects. The prior tactile knowledge improves the successful rate of 50% over traditional vision-based strategies.

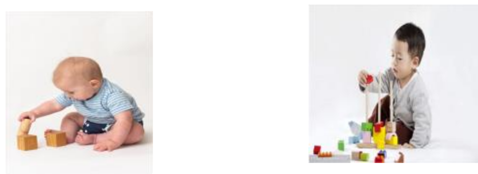
I. INTRODUCTION AND RELATED WORK

With the development of robotics and multi-modal technology, robots are able to acquire more and more human-like perception and abilities. However, they are still far from resembling the way humans manipulate objects, especially for stable grasping.

To better solve this problem, tactile sensing is introduced in grasping process. Di Guo [1], Emil Hyttinen[2], MM.Zhang [3], Deen Cockburn [4] and Hogan F R [5] et al. employ tactile data into current grasp to evaluate hypothetical configurations and guide the robot to regrasp by combining the stability estimation method and the grasp adaption method.

However, these methods depend on regrasping for many times, which no doubt waste much time to regulate and evaluate, until satisfying the stability required.

Let's imagine how humans grasp objects. Figure 1 (a) depicts an example where an infant tries to grasp toy blocks. It's obvious to find that he has no idea of where to grasp, and the certain result is falling or slippage. He has to regrasp for many times to regulate gradually. Meanwhile, Figure 1 (b) shows a grown-up boy can easily build great works with blocks. The reason for that is infants cannot justify which point is appropriate to grasp at the moment of looking at target objects for lack of massive tactile experience, which can be seen as a kind of prior tactile knowledge.



(a) (b)

Fig. 1. Illustration of stable grasp at different stage of humans

Therefore, we think that the regrasp policy delays robots staying at the infant stage, and what we want to do is to endow robots with prior tactile knowledge to act as a grown-up boy. To achieving this target, we propose a novel grasp mode, Bayesian Grasp, which applies much prior knowledge into the final grasp process, and we focus on prior tactile knowledge mainly in this paper.

The contributions of this paper can be concluded as follows:

1) Prior Tactile Knowledge Learning Framework. We propose a grasp quality Metric and a self-supervised tactile knowledge learning approach.

2) Bayesian Grasp Policy. A multi-phases architecture is proposed for stable grasp, which can generate good grasp configurations through a single RGB image.

II. PRIOR TACTILE KNOWLEDGE LEARNING

In this section, we build a novel tactile-based grasp quality metric. Then, we use this metric to evaluate grasp quality of each grasp.

A. Grasp quality metric

The grasp quality is evaluated through a continuous score from 0 to 1. Fig. 2. Shows tactile images and quality scores of three different grasping conditions.

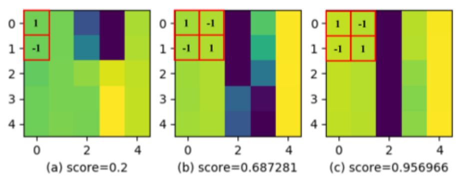


Fig. 2. Tactile images under 3 different stability conditions

To assess the quality of a grasp, we design a group of robotic shake actions: 2 fast end-effector move and 2 fast joint rotation. The grasp quality metric is:

*Resrach supported by NaturalScience Foundation of China (51775332, 51675329) Foundation.

Teng Xue, Wenhai Liu, Mingshuo Han, Jin Ma, Quanquan Shao and Weiming Wang are with the Institute of Knowledge Based Engineering, School of Mechanical Engineering, Shanghai Jiao Tong University, Shanghai, 200240, China(corresponding author phone: 008621-34206501; fax: 008621-34206840; e-mail: wangweiming@ sjtu.edu.cn, {xueteng, sjtu-wenhai, mrjay101, pan13855965084, heltonma, shao124}@sjtu.edu.cn).

$$
\begin{cases}
0 & \text{If grasp fails,} \\
0.5 \times \dfrac{i}{5} & \text{If falling occurs,} \\
0.5 \times e^{-x/3000} + 0.5 & \text{If no falling is detected,}
\end{cases}
\qquad (1)
$$

where i means the row number where the maximum value occurs after 2*1 kernel(Fig. 2 (a)) convolution operation, and x is the results after 2*2 kernel(Fig. 2 (b) (c)) convolution with tactile image.

B. *Prior tactile knowledge learning*

We define the stability status as four categories: failure, falling, slippery and stable, with the score 0, (0, 0.5], (0.5, 0.85] and (0.85, 1]. Five objects (Table I) are grasped 300 times to produce dataset.

Performance on known objects. We obtain a training accuracy of 92% and a testing accuracy of 86%. This result shows that given a grasp about a known object, we can get a stable grasp configuration reliably.

Performance on unknown objects. We use cross-validation strategy to evaluate the performance. The classification accuracy for each object is 46% (screw), 50% (yogurt), 85% (tennis container), 81% (flashlight) and 75% (metal bar). The results show that this method has the ability to generalize to other new objects.

what tactile sensor?

III. Bayesian Grasp

We propose Bayesian Grasp achitecture in this section, and the whole process can be divided into two phases: grasp region estimation and stable configuration generation.

A. *Grasp region estimation phase*

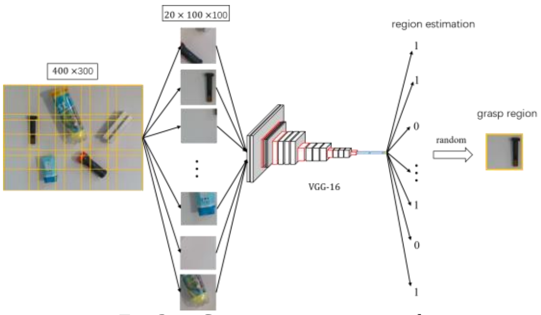


Fig. 3. Grasp region estimation phase

As Fig. 3 shows, we extract region candidates from the original input image with (400, 300) pixels by sliding window with (100,100) pixels, and 20 region candidates are generated, which will be classified as '0' and'1' through the network. One candidate will be determined as the grasp region from the potential graspable regions randomly.

B. *Stable configuration generation phase*

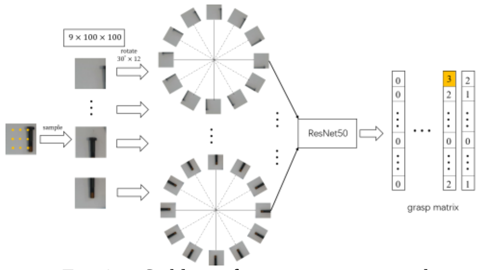


Fig. 4. Stable configuration generation phase

As Fig. 4 shows, in the (100, 100) pixels candidate determined before, 9 (3, 3) potential grasp points are extracted, centering at 9 new (100, 100) pixels candidates. To acquire proper grasp angle, those candidates are rotated every 30 degree, which means 108 (9×12) images are generated totally. All the images are fed into ResNet-50, and 108 classifications are outputted. Candidates labelled as '3' are determined as stable point and angle.

IV. Experiments

To contrast the performance of our proposed framework, we retrain a ResNet-50 model with new labels acquired only by visual information. The successful and unsuccessful grasps are labelled as '1' and '0'respectively after shaking. The rates of successful grasps (score>0.85) are listed in Table I. We can find that our method increases the overall accuracy by an average of approximately 25%.

V. Discussion and Conclusion

This paper proposes a novel framework to improve grasp stability by using prior tactile knowledge.

Discussion on experiment results. This method can be generalized within a certain range. And Table I shows that this method shows great improvement to heavy and out-of-shape objects.

Future work. Analyzing and Learning the changes among continuous tactile information would better improve the stability during manipulation process. Besides, designing a flexible tactile sensor with high resolution might be better to reflect the contact situation in detail. What's more, a set of tactile standard on acquisition, transmission and processing need to be built up, and the study of computer vision might be referred.

TABLE I. Prior Tactile Knowledge Improves Grasping Accuracy

Objects	screw	yogurt	tennis container	flashlight	metal bar
Images					
Vison-based Grasp	35%	43%	71%	65%	52%
Bayesian Grasp	75%	62%	93%	89%	83%
Relative improvement	114%	44%	31%	24%	60%