Proceedings of the 2nd WRC Symposium
on Advanced Robotics and Automation 2019
Beijing, China, August 21, 2019

Vision-based Adjustment with Toleration of Error before Grasping

Yifan Duan[1],Ting Lei[1], Peichen Wu[1], Nan Lin[1], Wenbo Chen[1], Xiaoping Chen[1]

***Abstract*— Automatic grasp has always been a fundamental topic in the field of robotics research. Only with the stable and high success rate grasp, can the robots behave well in some industrial scenes and man-machine interaction scenes which usually contain grasping tasks. However, the error of various devices caused by the diversity of the complicated environments mentioned above is always ignored by the traditional grasp frame assuming that the sensor and the actuator are precise. In this paper, we take error tolerance as the core idea, proposing a novel frame of grasp which includes global search and local adjustment. In addition to an under-actuated gripper, we also use a depth camera and an ordinary micro camera in the whole grasp system. Based on the data obtained by the cameras and an output from ResNet, the overall state of the gripper and the target object is estimated to make appropriate adjustments. This improvement enables the grasp system to adapt to more various unstructured environments. Finally, we conducted experiments on three common application scenarios proving the validity of the improvement.**

I. INTRODUCTION

Considering the behavior of human beings when grasping, humans obtain the position of the object with the help of the eyes firstly. Put the hand in the neighborhood of the object, and then adjust the hand to an optimal position by virtue of the experience. At the end, complete a successful grasp.

Similarly, the traditional automatic grasp problem is usually divided into two sequential phases: grasp selection and grasp execution. The distinction between it and the process of human's grasp is the lack of adjustment according to the experience before closing gripper. In order to make up for this shortcoming, the usual solution is precise control assuming that every part of the robot is accurate and errorless. For many cases, this assumption is tenable. However, in some complex and unstructured scenarios, the error is often not as negligible as we imagined. Sometimes, centimeter-level errors can be a trigger for a bad consequence that we don't want to see.

Some examples of the problems caused by the lack of necessary adjustments are as follows. For example, in some complex environments, the sensors and the actuators may not be as exact as usual because of the environmental factor such as the poor light, shake (in the factories) and etc. Some terrible consequences may occur with a higher probability than usual if the grasp is failed. The harm to the object and the harm to the environment caused by the failure is both not what we wish to see. What's more, when robots execute the man-machine interaction task, the instability of human's control causes the object to move from its original position, which also leads to the failure ultimately. If the controller can make targeted adjustments based on the real-time conditions of the object, most of the failures mentioned above can be avoided.

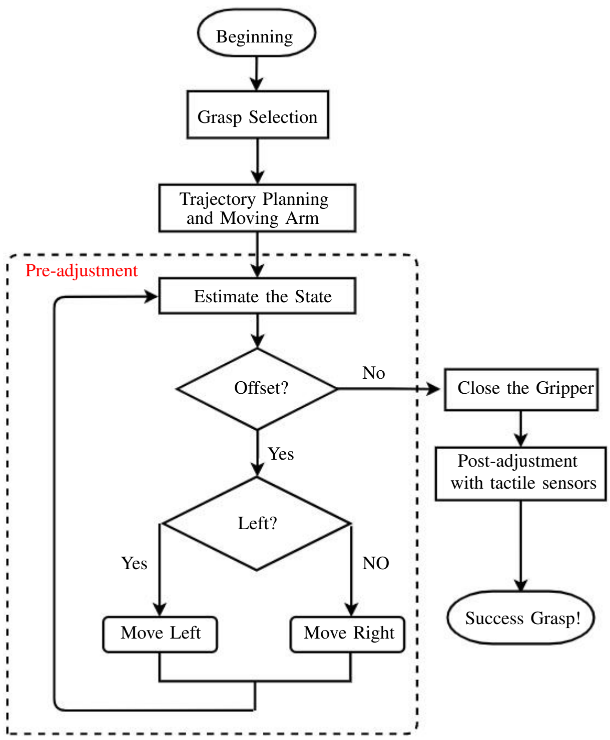


Fig. 1. Our updated grasp frame. The traditional grasp frame is outside the dashed box with regular grasp selection and grasp execution and the pre-adjustment step is in the dashed box including estimating the state and adjusting the gripper. We will discuss them detailedly in Section IV.

Some researchers are also studying the adjustments. These adjustments usually occur after grasping at least once to acquire contact information, which we call it post-adjustment. This adjustment is to judge the state of the object after the grasp is executed. However, this kind of adjustment isn't a substitute for the pre-adjustment which appends before grasping. Firstly, the sufficient condition of the post-adjustment is having touched. However, if there is a large offset between the gripper and the object at the beginning, the post-adjustment will not work at all. What's more, post-touch adjustment means that the irretrievable accident during the contact process are inevitable.

*This work is supported by the National Natural Science Foundation of China under grant U1613216 and 61573333.

[1] Yifan Duan, Ting Lei, PeiChen Wu, Nan Lin, Wenbo Chen and Xiaoping Chen are with the School of Computer Science and Technology, University of Science and Technology of China, Hefei, 230026, China. xpchen@ustc.edu.cn and dyf0202@mail.ustc.edu.cn

In this paper, we split the traditional execution into three parts: moving arm depending on the global grasp selection, pre-adjustment and closing the gripper. In Fig. 1, the updated grasp frame can be seen. To obtain the relative position of the gripper and the target, we fix a micro camera on the wrist of the mechanical arm, establishing a communication bridge between them. After the optimization, the failure due to equipment error is avoided and the wonderful passive adaptability of the under-actuated gripper is utilized. The obvious effects can be seen in the experimental section.

The research on related work is shown in Section II. The grasp selection algorithm is exposed in Section III. Section IV describes how the pre-adjustment is implemented. What' s more, the experiments can be found in Section V. At the end, Section VI is about the conclusion and the future work.

II. Related Work

The research on related work focused on three fields. At the beginning, we investigated the recent work on grasp selection with visual information. Then, we focused on other researchers' attempts on adjustment, most of which was post-adjustment and used feedback from tactile sensors. At last, we paid attention to how people evaluate the idea of fault tolerance in the robotics field.

Most of information perceived by human beings comes from vision, which many robotic grasp experiments are based on. One class of approaches uses a sliding window to detect regions of an RGBD image or a height map where a grasp is likely to succeed [1] [2]. Another method of grasp position selection is data driven. [3] describes a learning-based hand-eye coordination method for robots to grab from monocular images. The paper uses eight robotic hands to collect data and train the CNN to evaluate the motion of the gripper. It decides whether to grasp and how to grasp based on the prediction success rate. Such a single source of information makes it too expensive to get a similar success rate to other researchers. [4] and [5] both model common objects and then compare the new objects with the existing models for grabbing. In [6] and [7], the neural network is used to process the acquired image, but it still requires a lot of manpower to process the data. Our grasp selection is similar to [8] which uses the geometric information of the object to determine, but we extend the selection criteria in [8] to make it more suitable for our under-actuated gripper.

There are also many studies on adjustment, such as [9] [10] [11] [12] [13]. However, most of them assume that there is no error in the movement of the arm and the gripper has touched the object. [14] processes the measurements of the fingertip pressure array from the fixture and the manual accelerometer in real time to generate a robotic tactile signal to determine the gripper motion. [15] and [16] use CNN to learn to adjust the gripper with tactile information. However, due to the simple structure of the grippers, the objects that can be grasped and the applicable scenes are very limited. [17] establishes a database to guide how to adjust for different situations, but this method requires a lot of experiments to build the database. In [18], the under-actuated gripper and the tactile sensor are also used, but it can only work after the gripper and the object are in contact. In addition to the use of tactile sensors, [19] also takes into account some of the physical properties of the surface of the object. [20] and [21] use a tactile sensor to detect whether the object has slipped. Although it's not the same as our focus, but it provides inspiration for our next work. [22] uses a combination of visual and tactile information, and the overall idea is similar to ours. Leverage vision to observe scene-level information firstly, then learn local details necessary for task completion.

In this paper, we propose a method to adjust with vision so that the gripper can be as close as possible to the ideal situation before closing. Moreover, we do not calculate the error accurately, but aim to grab the object successfully. This task-oriented thinking is simple and effective, and the mathematical idea for quantization error is not suitable for automatic grasp. This fault tolerance is the minimization of the robot's "fault tolerance" mentioned in [23]. Researchers have shown that the integration of fault-tolerant technology into robotic systems improves reliability [24].

III. Grasp Selection

The selection of the grasp position is the foundation of the automatic grasp action. We use a depth camera (Kinect) to obtain the point cloud data of the target object as input to the grasp selection algorithm, while an appropriate positon for grasp is the output of the algorithm. This algorithm is a simplified version compared to our previous work [25]. Since we were paying attention to the grasp in an open environment without any obstruction, we took all the possible positions into account. This time, we focus on the complex environment, so we only make a more detailed calculation for one specific direction.

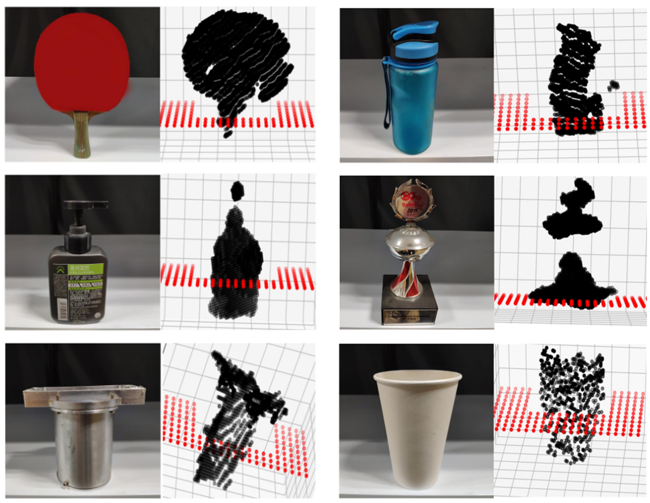


Fig. 2. Examples of the results (right) calculated by the grasp selection algorithm for the test objects (left). The black part in the right pictures is the point cloud of the targets and the red part represents the gripper. The test objects in turn is: a table tennis bat, a water bottle, a facial cleanser bottle, a trophy, a T-shaped iron casting and a paper cup.

The core idea is to take full advantage of the under-actuated gripper. The first question to consider is how to increase the contact area between the object and the gripper, so a smooth surface is a wonderful choice. In addition, the strongest feature of the gripper is its passive adaptability. Then, if the gripper can fit the shape of the object at the grasp position as much as possible, it will behave better. Finally, based on physics knowledge, the grasp position near the centre of gravity of the object will make the grasp more stable.

Based on the above ideas, we set up three parameters for each position, called “SMOOTHNESS”, “SHAPE”, “RELATIVE POSITION”. See details below.

A. Smoothness

We use SMOOTHNESS to describe the concave and convex condition of the surface. The larger the value of SMOOTHNESS is, the smoother the object is. To get the value of it, we calculate the widths of a position along the grasp direction, and then fit the widths to the form of a Fourier series:

$$\begin{aligned} f(x) =& a_0 + a_1 cos(xk) + b_1 sin(xk) + \\ & a_2 cos(2xk) + b_2 sin(2xk) + \\ & a_3 cos(3xk) + b_3 sin(3xk) ... \end{aligned} \tag{1}$$

Analyze the second derivative of the function. The surface fluctuation is described by the magnitude of the second derivative. A constant S describing the SMOOTHNESS can be calculated by the following formula.

$$s = 1 - (m/t) \tag{2}$$

Where m denotes the maximum of the second derivative and t denotes a threshold which depends on the gripper. Obviously, when the object is smooth following the definition, $s > 0$; but if we find the $s < 0$, it means that the object is not smooth.

B. Shape

Since we only need to know a rough shape, we fit the widths mentioned above into a quadratic function which is relatively simple to see the variation tendency of a specified interval (the front end of the object to the fingertip of the gripper which is assumed to hold the object). There are five situations in total:

- Rectangular (parallel on both sides)
- Trapezoid 1 (narrow front end and wide back end)
- Trapezoidal 2 (wide front end and narrow back end)
- Round (convex middle)
- Annulus (concave middle)

There is no denying that a better option is like the trapezoidal2 and round. So, the relationship between the maximum value of the quadratic function and the interval mentioned above is considered. If the maximum value is within the interval, a constant describing SHAPE can be calculated by the following formula, otherwise the constant is set to 0.

$$c = 1 - 4(m/d - 0.5)^2 \tag{3}$$

Where d denotes width between two fingertips after the gripper fully opens. Obviously, if the maximum is about half of the width, the value of c is the largest. If the maximum is too large, c will be negative.

C. Relative Position

RELATIVE POSITION is used to describe the offset between the grasp position and the centre of gravity. Since it isn't an easy thing to obtain the centre of gravity from the point cloud data directly, we use the center of the boundingbox of the object instead.

Using the following formula, a constant h is gotten.

$$h = 1 - (Z - Z_{center})/(Z_{max} - Z_{center}) \tag{4}$$

So far, we have acquired all the information we need: s (SMOOTHNESS), c (SHAPE), h (RELATIVE POSITION). We use the formula $w = \alpha s + \beta c + \gamma h$ to assign the three parameters with different weights to sort the position. The first one is the optimal grasp position obtained by this algorithm. Some examples are shown in Fig. 2

IV. ADJUSTMENT

A. The Idea and Its Significance

As mentioned earlier, pre-adjustment plays an important role in the entire grasp action. In order to enable the controller to obtain the real-time state of the object in the process of approaching the object, we install an ordinary micro camera at the wrist of the robot. Without the help of the global depth camera, the image obtained by the micro camera can feed back the overall state of the object and the gripper to the control program in real time so that the robot can adjust the gripper to an ideal position as much as possible before the gripper closing. We use a three-state finite-state machine (FSM) to describe this process in Fig.3. This FSM will guide the gripper to the proper position as long as the object appears within the field of view of the camera.

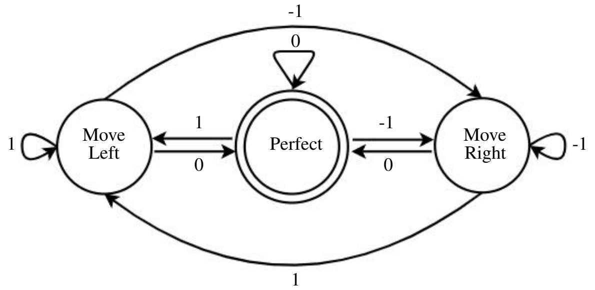


Fig. 3. The FSM. The concentric circles means the accepted state, while the other two circles means the corresponding adjustment method to different conditions. “-1” means the object is on the left side of the gripper, “1” means the object is on the right side and “0” means the gripper is in an ideal state for grasping. The combination of the FSM and the robot can be seen in the video together with this paper.

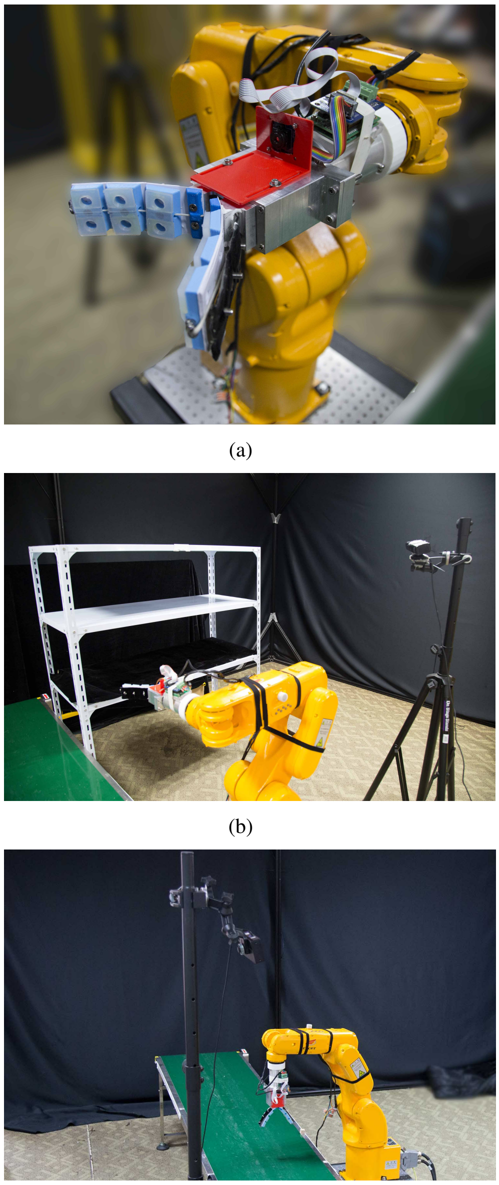
Fig. 4. The objects of the training set.

B. Introduction of The Model

In order to make the controller judge the state clearly after receiving the picture from the micro camera, we train the data collected with ResNet which was presented by Kaiming He in [26] in 2015. ResNet is arguably the most groundbreaking work in the field of computer vision and deep learning over the past few years. ResNet makes it possible to train hundreds or even thousands of layers, and still exhibits superior performance in this case. The core idea of ResNet is using the simple concept of residual learning to overcome the challenge of learning an identity mapping. A standard feed-forward CNN is modified to incorporate skip connections that bypass a few layers at a time. We encourage the readers to see [26] for more details on the ResNet architecture. In this paper, we use ResNet-50, a fifty layer deep residual model, to train our data.

C. Data Set Introduction and The Accuracy Rate

We collected about 3600 sets of data for six representative objects which are common in life shown in Fig. 4. A measurement system [27] is used to label the data, which is full automatic. The system uses the global cameras to capture the mark point and get their coordinate values which represent the position of the object and the gripper accurately. On the test set of these six objects, the accuracy reaches 93%. The model also has good generalization ability and accuracy rate is 87% when testing other items that are not in the training set.



(a)

(b)

(c)

Fig. 5. The experimental environment. (a) The under-actuated and the mechanical arm. The micro camera is in the middle of the red part. (b) The depth camera is placed on the right and the target objects will be placed on the shelf. (c) The global camera is above the conveyor belt to locate the starting position of the object.

V. EXPERIMENT

First of all, let me show the experimental equipment. The soft under-actuated gripper is designed in our previous work [28] and the mechanical arm is produced by EFORT, a company in China [29]. The robot is shown in Fig. 5(a).

A. Tolerate The Error

At the beginning, we tested the performance of the optimized grasp frame. In order to simulate a real unstructured environment, we place the Kinect far away from the object, only getting a small part of the target's point cloud. The experimental environment can be seen in Fig. 5(b).

To simulate the error of devices in the real environment, we shift the object slightly during the movement of the arm. After testing, our gripper can be adjusted to the corresponding position flexibly when the offset is within 5 cm.

We conducted a series of comparative experiment on 5 objects without pre-adjustment and with pre-adjustment. Each object tried to be caught for 6 times for different circumstances and the results are shown in Table I. The judgment criteria for successful grasp is shaking the gripper slightly after closing and lifting and see if the object is still held by the gripper. Obviously, the method with pre-

TABLE I
DETAILS OF GRASPING OBJECTS WITH AND WITHOUT PRE-ADJUSTMENT FACING DIFFERENT OFFSET.

Object	Without Adjustment						With Adjustment					
	0cm	1cm	2cm	3cm	4cm	5cm	0cm	1cm	2cm	3cm	4cm	5cm
Bottle	Y	Y	N	N	N	N	Y	Y	Y	Y	Y	Y
Paper cup	Y	Y	Y	N	N	N	Y	Y	Y	Y	Y	Y
Box	Y	N	N	N	N	N	Y	Y	Y	Y	Y	Y
Watering can	Y	N	N	N	N	N	Y	Y	Y	Y	Y	Y
Mug	Y	Y	N	N	N	N	Y	Y	Y	Y	Y	Y

adjustment ignoring the offset had an excellent performance.

B. Grasp The Object On The Conveyor Belt

Conveyor belts are widely used in logistics and warehouses. Since our adjustment step allows the gripper to follow the object, we make a bold trial about the grasp mission on the conveyor belt.

The traditional method to detect objects is using the laser sensor, which means that it requires a lot of debugging time before use to make the running speed of the object and the speed of the gripper stay tuned. However, our method is a plug and play, almost no debugging.

First of all, a global camera is used to obtain the relative position of the object on the conveyor. The mechanical arm will move the gripper to a position that is in line with the object when waiting the object to enter the visual range of the camera on the wrist.The experimental environment can be seen in Fig.5(c).

Since the object always appear on one specific side of visual range (assuming left), the hand will be adjusted to the left according to the ResNet, showing an opposite trend with the object. When the camera recognizes a suitable position, the gripper will close decisively. However, if the controller misses the execution signal, the object will appear on the other side of the visual range (assuming right). For our purpose, we set the arm's translation speed to a value slightly faster than the conveyor speed, so the gripper will always catch up with the object and perform a stable grasp.

C. Man-machine Interactive Experiment

To prove the effectiveness of our algorithm in daily life, we conducted experiment about fetching objects from people. The input to our grasp selection algorithm is composed of the point cloud of both people's hand and the object held in the hand. However, according to our grasp selection algorithm, the position where the hand is in will never be selected because of the rough surface of the finger.

What's more, one problem that cannot be ignored is that human's control to the hand is not as stable as the mechanical arm's control to the gripper. It means that the object is likely to have left its original position when the robot is trying to reach the position having been calculated by the algorithm. Here, our adjustment step play an important role so that the gripper can always find the moved object or even follow the movement of the human's hand and finally achieve a successful handover.

At the end, all the experiment can be seen in the video together with this paper.

VI. CONCLUSION

In this paper, we insert the pre-adjustment step into the traditional grasp process. With the feedback about the real-time information of the target object from the micro camera, the controller can make corresponding adjustment according to the actual state of the object. The adjustment step reduces the requirements for the accuracy of various devices in the traditional grasp, and can also cope with many complicated tasks. We conducted three experiments, proving the effectiveness and the practicality of our improved algorithm. With the development of technology, robots have gradually appeared in more and more scenes in our lives, while the tasks faced are increasingly diverse and complex, so the pre-adjustment that can minimize the possibility of failure and handle multiple tasks will be an essential part of the automatic grasp.

In the future, on one hand, we intend to try to extract more effective information from the image taken by the micro camera, such as detecting slip, detecting the stability after grasp and etc. On the other hand, multi-sensor fusion is also our research direction. How to integrate the data obtained from different sensors and extract more information, will be pay attention to continuously.

REFERENCES

[1] D. Fischinger and M. Vincze, "Empty the basket-a shape based learning approach for grasping piles of unknown objects," in *2012 IEEE/RSJ International Conference on Intelligent Robots and Systems*. IEEE, 2012, pp. 2051–2057.

[2] D. Fischinger, M. Vincze, and Y. Jiang, "Learning grasps for unknown objects in cluttered scenes," in *2013 IEEE International Conference on Robotics and Automation*. IEEE, 2013, pp. 609–616.

[3] S. Levine, P. Pastor, A. Krizhevsky, J. Ibarz, and D. Quillen, "Learning hand-eye coordination for robotic grasping with deep learning and large-scale data collection," *The International Journal of Robotics Research*, vol. 37, no. 4-5, pp. 421–436, 2018.

[4] C. Goldfeder and P. K. Allen, "Data-driven grasping," *Autonomous Robots*, vol. 31, no. 1, pp. 1–20, 2011.

[5] D. Kappler, J. Bohg, and S. Schaal, "Leveraging big data for grasp planning," in *2015 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2015, pp. 4304–4311.

[6] I. Lenz, H. Lee, and A. Saxena, "Deep learning for detecting robotic grasps," *The International Journal of Robotics Research*, vol. 34, no. 4-5, pp. 705–724, 2015.

[7] E. Johns, S. Leutenegger, and A. J. Davison, “Deep learning a grasp function for grasping under gripper pose uncertainty,” in *2016 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2016, pp. 4461–4468.

[8] A. ten Pas and R. Platt, “Using geometry to detect grasp poses in 3d point clouds,” in *Robotics Research*. Springer, 2018, pp. 307–324.

[9] H. Dang and P. K. Allen, “Stable grasping under pose uncertainty using tactile feedback,” *Autonomous Robots*, vol. 36, no. 4, pp. 309–330, 2014.

[10] Y. Bekiroglu, J. Laaksonen, J. A. Jørgensen, V. Kyrki, and D. Kragic, “Assessing grasp stability based on learning and haptic data,” *IEEE Transactions on Robotics*, vol. 27, no. 3, p. 616, 2011.

[11] J. Schill, J. Laaksonen, M. Przybylski, V. Kyrki, T. Asfour, and R. Dillmann, “Learning continuous grasp stability for a humanoid robot hand based on tactile sensing,” in *2012 4th IEEE RAS & EMBS International Conference on Biomedical Robotics and Biomechatronics (BioRob)*. IEEE, 2012, pp. 1901–1906.

[12] D. Cockbum, J.-P. Roberge, A. Maslyczyk, V. Duchaine *et al.*, “Grasp stability assessment through unsupervised feature learning of tactile images,” in *2017 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2017, pp. 2238–2244.

[13] R. Calandra, A. Owens, M. Upadhyaya, W. Yuan, J. Lin, E. H. Adelson, and S. Levine, “The feeling of success: Does touch sensing help predict grasp outcomes?” *arXiv preprint arXiv:1710.05512*, 2017.

[14] J. M. Romano, K. Hsiao, G. Niemeyer, S. Chitta, and K. J. Kuchenbecker, “Human-inspired robotic grasp control with tactile sensing,” *IEEE Transactions on Robotics*, vol. 27, no. 6, pp. 1067–1079, 2011.

[15] R. Calandra, A. Owens, D. Jayaraman, J. Lin, W. Yuan, J. Malik, E. H. Adelson, and S. Levine, “More than a feeling: Learning to grasp and regrasp using vision and touch,” *IEEE Robotics and Automation Letters*, vol. 3, no. 4, pp. 3300–3307, 2018.

[16] B. S. Zapata-Impata, P. Gil, and F. Torres, “Non-matrix tactile sensors: How can be exploited their local connectivity for predicting grasp stability?” *arXiv preprint arXiv:1809.05551*, 2018.

[17] H. Dang and P. K. Allen, “Grasp adjustment on novel objects using tactile experience from similar local geometry,” in *2013 IEEE/RSJ International Conference on Intelligent Robots and Systems*. IEEE, 2013, pp. 4007–4012.

[18] M. Liarokapis and A. M. Dollar, “Learning the post-contact reconfiguration of the hand object system for adaptive grasping mechanisms,” in *2017 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2017, pp. 293–299.

[19] M. Li, Y. Bekiroglu, D. Kragic, and A. Billard, “Learning of grasp adaptation through experience and tactile sensing,” in *2014 IEEE/RSJ International Conference on Intelligent Robots and Systems*. Ieee, 2014, pp. 3339–3346.

[20] F. Veiga, H. Van Hoof, J. Peters, and T. Hermans, “Stabilizing novel objects by learning to predict tactile slip,” in *2015 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2015, pp. 5065–5072.

[21] J. M. Romano, K. Hsiao, G. Niemeyer, S. Chitta, and K. J. Kuchenbecker, “Human-inspired robotic grasp control with tactile sensing,” *IEEE Transactions on Robotics*, vol. 27, no. 6, pp. 1067–1079, 2011.

[22] M. M. Zhang, R. Detry, M. Larry, and D. Kostas, “Tactile-vision integration for task-compatible fine-part manipulation,” in *Robotics: Science and Systems (RSS) Workshop on Revisiting Contact C Turning a Problem into a Solution*, 2017.

[23] M. L. Visinsky, J. R. Cavallaro, and I. D. Walker, “Robotic fault detection and fault tolerance: A survey,” *Reliability Engineering & System Safety*, vol. 46, no. 2, pp. 139–158, 1994.

[24] D. L. Hamilton, I. D. Walker, and J. K. Bennett, “Fault tolerance versus performance metrics for robot systems,” *Reliability Engineering & System Safety*, vol. 53, no. 3, pp. 309–318, 1996.

[25] P. Wu, N. Lin, Y. Duan, T. Lei, L. Chai, and X. Chen, “An automatic grasp system with sensor feedback based on soft gripper,” in *2018 WRC Symposium on Advanced Robotics and Automation (WRC SARA)*. IEEE, 2018, pp. 1–7.

[26] K. He, X. Zhang, S. Ren, and J. Sun, “Deep residual learning for image recognition,” in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770–778.

[27] Calibration of optitrack mocap. [Online]. Available: https://v20.wiki.optitrack.com/index.php?title=Calibration

[28] N. Lin, P. Wu, X. Tan, J. Zhu, Z. Guo, X. Qi, and X. Chen, “Design and analysis of a novel sucked-type underactuated hand with multiple grasping modes,” in *International Conference on Robot Intelligence Technology and Applications*. Springer, 2017, pp. 299–312.

[29] Products of efort. [Online]. Available: http://www.efort.com.cn/product_category.php