Estimating End-Effector 3D Position using a Single Monocular Microscopic Image for Robotic Micromanipulation

Jiaqi Wang, Jiaqi Chen, Zhuoran Zhang

Abstract— **Visual servo control of end-effectors is a crucial step in robotic micromanipulation. In the three-dimensional positioning problem of end-effector, while methods have been developed for visually detecting the x-axis and y-axis positions of the end-effector tip, it remains challenging to obtain visual feedback of the z-axis positions. In this paper, a new strategy is proposed to estimate the z-axis position of the end-effector. Instead of using depth-from-focus and depth-from-defocus methods, we transform z-axis positioning problem into a multiclass classification problem. Our strategy takes a monocular image of the end-effector as input, classifies it into different depth intervals, and outputs the focal plane z-axis position for that interval. A deep learning model is developed to solve the multiclass classification problem. Considering of the shallow depth of field of an optical microscope, a novel loss function is proposed to penalize misclassification. Using glass micropipettes as an example, the deep learning model achieves an accuracy of 96.1% for depth prediction/classification. The proposed strategy provides a new method for locating the out-of-focus depth of the end-effector and for providing 3D visual feedback for robotic micromanipulation.**

Keywords: End-effector manipulation, Automation at micro-scale, Robot Vision

I. Introduction

The past decades have witnessed significant development of robotic micromanipulation techniques. Under the visual guidance of an optical microscope, an end-effector is automatically controlled for the assembly of microparts, material characterization, and manipulation of biological cells [1]. In robotic micromanipulation, obtaining the visual feedback of end-effector position is an essential step in visual servo control. The positioning of the end-effector is performed in three-dimensional space, thus requiring the robot system to have three-dimensional (x-y-z) perception of the location of the end-effector.

Within the focal plane (x-y plane), the object being imaged (i.e., end-effector) is clearly visible, and the visual detection algorithms for its position have been relatively mature. Outside the focal plane, the end-effector is blurred, and it is difficult to obtain the depth information in z-direction (Fig. 1). The use of algorithms to achieve depth prediction in monocular vision would be significantly more convenient in the field of micromanipulation compared to adding additional hardware support. In monocular microscopic vision, z-direction position estimation algorithms are usually divided into two categories: depth from focus [2] and depth from defocus [3].

Different from existing depth-from-focus methods and depth-from-defocus methods, this paper utilizes the characteristics of microscope depth of field and transforms the continuous depth estimation problem into a depth classification problem. The multi-class depth classification problem is then solved by machine learning, which takes a single end-effector image as input and predicts/classifies its corresponding depth of the originating imaging plane. Using glass micropipettes as an example, the developed machine learning model achieved a prediction accuracy of 96.1%.

II. Solving Depth Prediction as a Multi-class Classification Problem

A. Depth Prediction of Micropipette under Microscope

The traditional methods cannot meet the demand of dynamic depth prediction of the micropipette under microscope in terms of both speed and accuracy, so a new method for estimating the out-of-focus depth of pipette is proposed in this paper.

The depth of field of an optical microscope is limited, and objects within the depth of field are simultaneously in focus. Hence, the continuous out-of-focus depth can be approximated by dividing it into multiple intervals, and the observed micropipette images in each interval show an approximate state of the focal plane with the same degree of out-of-focus as that interval. This could naturally transform the depth estimation problem into a classification problem: to which interval (originating imaging focal plane) does the micropipette image belong to. Therefore, the continuous depth prediction problem of the pipette under the microscope is transformed into a classification problem: classify the pipette image to different focal plane intervals, input an image, predict which focal plane interval it is in and output its Z-axis position (focal plane, see Fig.2). Then, the multi-class classification problem can be solved by using machine learning methods. In this paper, a classification method based on ResNet-34 is proposed and compared with other machine learning algorithms and existing traditional methods based on focus measure.

B. Loss Function Design with Penalty Coefficients

In order to improve the classification accuracy of the model, we add penalty coefficients to the cross-entropy loss function. The images in the pipette dataset collected based on depth gradient are strongly correlated, and there is a correspondence between the difference in depth gradient and the difference in category. The larger the category difference is, the larger the depth gradient difference is. To improve the classification accuracy of the model for similar depth images, we increase

The authors are with School of Science and Engineering, The Chinese University of Hong Kong, Shenzhen, 517182, Guangdong Province, China (e-mail: wangjiaqi @cuhk.edu.cn, 119010017@cuhk.edu.cn). Corresponding author: Zhuoran Zhang (zhangzhuoran@cuhk.edu.cn).

This work is supported by the University Development Fund of CUHKSZ (UDF01002141), National Natural Science Foundation of China under Grant (62203374), and Guangdong Basic and Applied Basic Research Foundation (2021A1515110023).

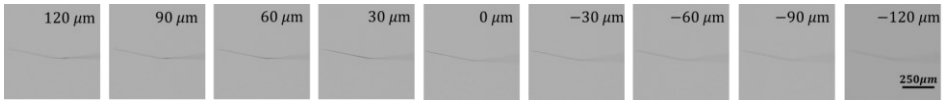

Figure 1. Images of the same pipette at different z-positions, marked on top of corresponding images.

the loss for the case of wrong prediction, which makes the algorithm pay more attention to it. The new loss function is designed as

$$WLoss = -\frac{1}{N}\sum_{n=1}^{N} W_n Log(P_n,\ i) \quad (1)$$

$$W_n = \left|{L_i}^2 - {L_n}^2\right| + 1 \quad (2)$$

where L_i is the ground truth, L_n is the predicted label value, and P_n is the predicted probability of SoftMax output.

C. Collection of the Training Dataset and Model Training

The robotic micromanipulation system consists of a standard inverted microscope, a CMOS camera and a motorized micromanipulator which could move the z-axis. The end-effector is a glass micropipette used for cell injection in clinical infertility treatment. A dataset consists of 900 images (100 z-stacks and each z-stack containing 9 images) of micropipette was collected to train the model.

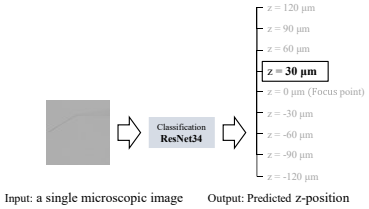

Figure 2. The proposed approach takes a single microscopic image of the end-effector as input and uses a ResNet34 model to predict its corresponding z-position.

III. Result and Discussion

A. Performance of the Depth Prediction Model

The depth prediction model was evaluated by its prediction accuracy, model size, training time and inference time. As summarized in Table 1, the model performance was compared with that of conventional deep learning models, including VGG, GoogLeNet. The highest accuracy is ResNet34 with WLoss with 96.1%, followed by ResNet34 with 80.2%. By adding penalty coefficients to the cross-entropy loss function, the F1-score is also improved by 16.2%. ResNet34 with WLoss also achieves the highest F1-score among these models. For inference speed (frames per second, FPS), typically, a frame rate of not lower than 30 FPS is regarded as real-time for potential robotic micromanipulation tasks [2]. Considering the trade-off between accuracy and inference time, ResNet-34 with WLoss was finally chosen as the depth prediction model.

Table 1. Performance of z-position predicted by different DNN-based models.

Model	*Accuracy*	*F1-score*	*Model Size*	*FPS*
ResNet-34	80.2%	80.1%	85.3 M	31
GoogLeNet	75.3%	74.9%	41.4M	32
AlexNet	61.7%	62.1%	58.4M	42
ResNet-34+WLoss	**96.1%**	**96.3%**	**85.3 M**	**30**

B. Comparison of Depth Estimation using the Prediction Model versus Conventional Focus Measure Methods

The traditional methods use different focus measure for in-focus image selection, we selected four more representative methods as baseline to compare with the machine learning based methods [4].

Results of the focus-measure methods are shown in Fig.3, where z = 0 μm is the ground truth value of the in-focus micropipette image. However, the four focus-measure methods all reached their peak focus-measure scores at -60 μm and -90 μm, resulting in an average absolute error of 75 μm. Obviously, the position of the corresponding micropipette tip is not in the in-focus position (see black dots in Fig.3). The recognition error of the traditional algorithm is between 50% and 75%.

The large error of focus-measure methods is mainly because adjusting the depth of the micropipette upward does not lead to the defocusing of the whole image, but to the movement of the focus point. These focus-measure methods can only determine whether the image is in-focus by the change of the pixel gradient. Since the in-focus interval of the end-effector will move from its tip to the body part when moving upward and does not disappear, the focus area becomes larger instead. Therefore, the pixel gradient still exists, so the focus-measure methods cannot accurately judge the focus situation of the tip, and will produce a large misidentification due to the shift of the in-focus area.

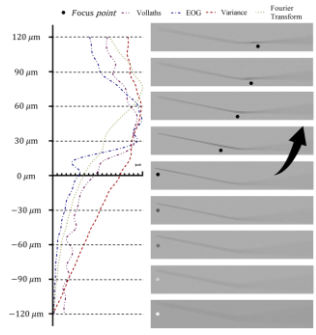

Figure 3. Performance of in-focus image selection and the actual pipette image corresponding to the Z-axis. The black dot indicates the position of the focus, and the arrow indicates the trend of the movement.

References

[1] Z. Zhang, X. Wang, J. Liu, C. Dai, and Y. Sun, "Robotic micromanipulation: Fundamentals and applications," Annu. Rev. Control Robot. Auton. Syst., vol. 2, no. 1, pp. 181–203, 2019

[2] J. Liu et al., "Automated vitrification of embryos: A robotics approach," IEEE Robot. Autom. Mag., vol. 22, no. 2, pp. 33–40, 2015.

[3] K. M. Taute, S. Gude, S. J. Tans, and T. S. Shimizu, "High-throughput 3D tracking of bacteria on a standard phase contrast microscope," Nat. Commun., vol. 6, no. 1, p. 8776, 2015.

[4] T. T. E. Yeo, S. H. Ong, Jayasooriah, and R. Sinniah, "Autofocusing for tissue microscopy," Image Vis. Comput., vol. 11, no. 10, pp. 629–639, 1993.