

EXHIBIT 1041



All



ADVANCED SEARCH

Conferences > 2019 Chinese Automation Congr... ?

Real Time and Robust 6D Pose Estimation of RGBD Data for Robotic Bin Picking

Publisher: IEEE

Cite This



PDF

<< Results

Linpeng Peng ; Yongsheng Zhao ; Shuailong Qu ; Yifeng Zhang ; Fang Weng All Authors ...

1
Cites in
Paper

448
Full
Text Views



Alerts

Manage Content Alerts
Add to Citation Alerts

Abstract

Document Sections

- I. Introduction
- II. Related Work
- III. Algorithm and Method
- IV. Experimental Results
- V. Conclusions

Authors

Figures

References

Citations

Keywords



Downl
PDF

Abstract:In this paper, we propose a real time and robust method to address the issues of 6D pose estimation for robot bin picking utilizing a low cost 3D sensor. RGB-D data have ... **View more**

► Metadata

Abstract:

In this paper, we propose a real time and robust method to address the issues of 6D pose estimation for robot bin picking utilizing a low cost 3D sensor. RGB-D data have been widely in industrial applications to deal with 6D pose estimation problems. Different from previous methods of registration-based with time-consuming and point cloud information only with poor accuracy, our approach uses a pinhole camera model and the geometric relationship to correlate the point cloud data and RGB pixels, which is faster and almost same accurate compared with registration based methods. Experimental results show that our method is computationally efficient and can reach an approximate centimeter accuracy. Furthermore, it is also much lower cost and easier to implement when Compared with anterior methods.

Published in: 2019 Chinese Automation Congress (CAC)

 Contents

SECTION I. Introduction



The scheme of 6D pose estimation has been widely studied in recent years. With different sorts of cheap RGB-D sensors invented, a lot of different styles of 3D sensors in the market that we can purchase. We can infer poses of objects by RGBD sensors even in poorly-lighted environments more accurately than only RGB image.

Our task is motivated by in industrial applications of logistics. We deal with the work of grasping objects of known express boxes and placing them to given locations. We intend to solve this task as low cost and can realize real time performance. Meanwhile, the pose accuracy should up to approximate centimeter.

For the sake of above destination, the original method is to segment point cloud and get different primitives. Then, computing primitives' position and rotation. For improving the precision of position, RGB information is also deployed as precision can up to millimeter by processing image according to repeated experiments. One of the examples is usual in 3D reconstruction [1]. Projecting 3D points onto image to get edge detection result can improve the accuracy a lot. Another example, Zhang et al. proposed a split-and-merge method to detect planar surfaces indoors based RGB-D data in [2]. Owing to mapping each segmented image region of depth image, their plane segmentation accuracy can arrive at a very fine result.

Hence, 3D point cloud projection with image is essential to promote 6D pose estimation accuracy. We propose a novel approach to solve the projection problem by pinhole camera model rather than region registration, which maps 2D RGB image to depth image (3D point cloud). This method is analyzed at length in Section III.

We make three contributions in this paper. (1) We propose a novel strategy to solve the problem of projection in RGB image point and 3D point cloud with the pinhole camera model. (2) We design a robotic bin picking system by integrating multiple setups for our algorithm and demonstrate real-time, precise and robust performance with the pose estimation compared with previous work [1, 2, 10]. (3) We propose the method to measure the pose estimation accuracy as a chessboard for reference. The remainder of this paper is as follows. Section II is reviewed by Related work, and Section III introduces the Algorithm and methods of our robotic bin picking system. Section IV shows the experimental steps and results of our approach. Finally, the conclusion is provided in Section V.

SECTION II. Related Work

Real Time and Robust 6D Pose Estimation of RGB-D Data for Robotic Bin Picking

Linpeng Peng
Polytechnic Institute
Zhejiang University
Hangzhou, China
*Email: penglinpeng@zju.edu.cn

Yongsheng Zhao
Binhai Industrial Technology Research
Institute
Zhejiang University
Tianjin, China
Email: zhaoyongsong@zju.edu.cn

Shuailong Qu
Polytechnic Institute
Zhejiang University
Hangzhou, China
Email: qushuailong@zju.edu.cn

Yifeng Zhang
Binhai Industrial Technology Research
Institute
Zhejiang University
Tianjin, China
Email: zhangyifeng@zjubh.com

Fang Weng
Binhai Industrial Technology Research
Institute
Zhejiang University
Tianjin, China
Email: wengfang@zjubh.com

Abstract—In this paper, we propose a real time and robust method to address the issues of 6D pose estimation for robot bin picking utilizing a low cost 3D sensor. RGB-D data have been widely in industrial applications to deal with 6D pose estimation problems. Different from previous methods of registration-based with time-consuming and point cloud information only with poor accuracy, our approach uses a pinhole camera model and the geometric relationship to correlate the point cloud data and RGB pixels, which is faster and almost same accurate compared with registration based methods. Experimental results show that our method is computationally efficient and can reach an approximate centimeter accuracy. Furthermore, it is also much lower cost and easier to implement when Compared with anterior methods.

Keywords—6D pose estimation, robot, bin picking, 3D point clouds, pinhole camera model

I. INTRODUCTION

The scheme of 6D pose estimation has been widely studied in recent years. With different sorts of cheap RGB-D sensors invented, a lot of different styles of 3D sensors in the market that we can purchase. We can infer poses of objects by RGB-D sensors even in poorly-lighted environments more accurately than only RGB image.

Our task is motivated by in industrial applications of logistics. We deal with the work of grasping objects of known express boxes and placing them to given locations. We intend to solve this task as low cost and can realize real time performance. Meanwhile, the pose accuracy should up to approximate centimeter.

For the sake of above destination, the original method is to segment point cloud and get different primitives. Then, computing primitives' position and rotation. For improving the precision of position, RGB information is also deployed as precision can up to millimeter by processing image according to repeated experiments. One of the examples is usual in 3D reconstruction [1]. Projecting 3D points onto image to get edge detection result can improve the accuracy a lot. Another example, Zhang et al. proposed a split-and-merge method to detect planar surfaces indoors based RGB-D data in [2]. Owing to mapping each segmented image region of depth image, their plane segmentation accuracy can arrive at a very fine result.

Hence, 3D point cloud projection with image is essential to promote 6D pose estimation accuracy. We propose a novel

approach to solve the projection problem by pinhole camera model rather than region registration, which maps 2D RGB image to depth image (3D point cloud). This method is analyzed at length in Section III.

We make three contributions in this paper. (1) We propose a novel strategy to solve the problem of projection in RGB image point and 3D point cloud with the pinhole camera model. (2) We design a robotic bin picking system by integrating multiple setups for our algorithm and demonstrate real-time, precise and robust performance with the pose estimation compared with previous work [1, 2, 10]. (3) We propose the method to measure the pose estimation accuracy as a chessboard for reference. The remainder of this paper is as follows. Section II is reviewed by Related work, and Section III introduces the Algorithm and methods of our robotic bin picking system. Section IV shows the experimental steps and results of our approach. Finally, the conclusion is provided in Section V.

II. RELATED WORK

The current methods for existing 6D object pose estimation consist generally of two pipelines. In this section, we review current approaches extending from traditional methods to modern popular deep learning based methods.

A. Pose from traditional methods

Traditional 6D pose estimation approaches are mainly feature-based, descriptor-based. Feature-based ways, are also the earliest object pose estimation method, which utilize different kinds of features, such as image edges, sharp corners and other more complex features to matches sparse feature points between 2D images and 3D object models [3]. Though superior in time-consuming works, these approaches show weak performance when in low-texture or cluttered background. These disadvantages also make them less used in practical applications.

For solving above problems, descriptor-based methods are proposed by scholars. In such cases, geometric descriptor and templates matching methods often work well. Papazov et al [4]. introduced a method which combined a robust geometric descriptor, a hashing technique with an efficient, localized RANSAC-simple sampling way. Although their method performed well on noisy, cluttered and unsegmented range scans, it could not up to real time demands. The Point Pair Feature, another descriptor approach, has been used extensively largely in 6D pose estimation method [5]. Choi et

al. [6] proposed a voting-based algorithm applicable to 3D sensors and constructed four Pair Features, which were Surface-to-Surface, Boundary-to-Boundary, Surface-to-Boundary, Line-to-Line. They tested the algorithm in a bin-picking system with a pickup success rate of over 98%, pose estimation error less than 0.3 mm for translations and 0.3° for rotations. Nevertheless, their industrial level RGB-D camera was costly in considerable numbers. One of recent work about templates matching methods [7], which utilized point tuple features for the quick generation of transformation guesses in a Random Sample Consensus (RANSAC) setting to carry on template matching. The improved strategy could reach 92.7% mean recall and only costed 4.9 seconds. However, the time cost of 4.9 seconds often could not satisfy industrial pace in robotic bin picking.

Other works also take advantage of different information to get pose information, such as the method of registration. In [8], Choi et al. presented edge-based pair-wise registration and a pose-graph SLAM method based on this registration, the result of experimental showed their approach had reached state-of-the-art. Chattopadhyay et al. [9] obtained smooth silhouette shape by register the depth and RGB frames from Kinect. Mapping edge or region point of RGB image to 3D point cloud is adopted in [2, 10]. Moreover, in [1], Chen et al. project 3D points onto image for getting edge detection result to improve the 3D reconstruction accuracy. Though the method of region projection improves the accuracy a lot, it is time-consuming to establish mapping relations in 3D points and image points by registration. Few of them can fulfil both accuracy and time consuming requirements. Even though there are schemes performing better in above detriments, it is always difficult to implement.

B. Pose from deep learning based methods

With deep learning technique developing, various neural network models have been designed to solve the limitations of traditional methods in computer vision. There are also more and more deep learning architectures that straightly perform 6D pose estimation of 3D computer vision.

Similar to the template-matching based methods, Kehl et al. introduced a new model SSD-6D [11], which was based on SSD archetype to detect 3D object instances and estimate corresponding 6D poses from RGB data. They used the thought of classifying to work out the difficulty of object pose estimation. Their approach in the paper exceeded current state-of-the-art methods that used RGBD data on multiple challenging datasets. However, this approach utilized the routine of bounding boxes in 2D instead of directly getting translations and rotations. Therefore, 6D pose estimation could not up to the-state-of-art precision which was demanded in high accuracy bin picking works. Actually, this method showed with poor performance in their experiments of some objects, such as ape and holep.

DenseFusion [12], one of the latest research result for 6D pose estimation by deep learning. Unlike previous works extract information from the RGB image and depth distinctly. This paper designed one technique to fuse geometric and appearance information in a heterogeneous architecture. Iterative refinement approach was used to further improve the pose estimation. Although the model performance surpassed the current sparkling works. Nonetheless, the model architecture is so complex that we have no enough time and GPU resources to train and run it.

Despite the fact that deep learning based methods for 6D pose estimation is a promising and challenging work in the future. Not only this way can deal with the cluttered backgrounds and occlusions of objects, but also arrive at the outstanding precision.

However, these methods need collect a large number of images of objects for training, which is time-consuming. Moreover, training a neural network model not only costs expensive computational resources, but also some industrial personal computers are not equipped with graphic card units (GPU), which are essential to train and run network model. Therefore, it is difficult to use these methods to work in a factory environment with poor computer configuration.

III. ALGORITHM AND METHOD

In our approach, point cloud and RGB image are processed in different style. Inspired by previous work [10], we use a Sobel operator to extract RGB edge and get box RGB pixel coordinate center. In order to find the corresponding point in IR camera coordinate of center pixel point, we project this center pixel to 3D space and get the corresponding point in IR camera coordinate. Random Sample Consensus (RANSAC) method is adopted to address point cloud plane segmentation problem and get box plane equation. Instead of projecting region of RGB image to depth image in [2] as an entire search throughout the full HD image will cause a lot of useless computational costs, we use pinhole camera model which is a physical model often used in camera calibration. In the end, we utilize this method to correlate RGB edge detection results with point cloud outcome for calculating 6D pose of different size boxes. Our approach is separated into three parts as shown in Fig. 1.

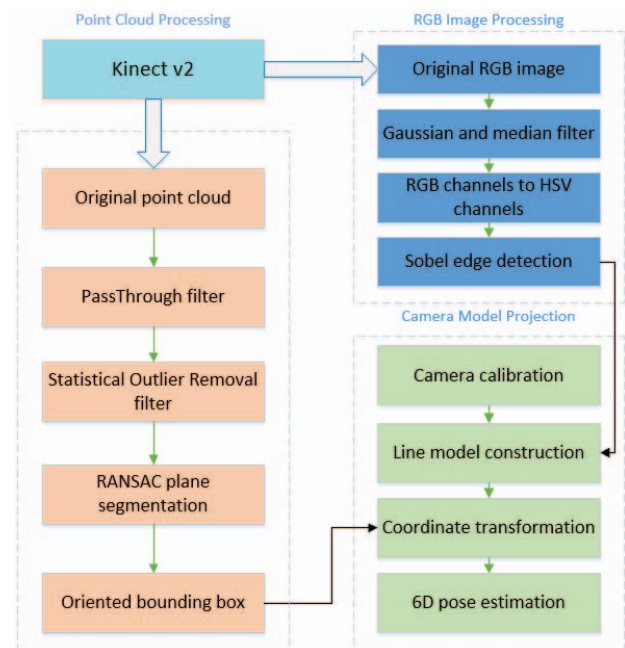


Fig. 1 Flow chart of the proposed method

A. Edge detection in the RGB image

In the preprocessing step, we use a Gaussian filter and Median filters to remove the influence of noise. Because the RGB channel is not robust in illumination tuning, we split RGB image to HSV channel [13] and set threshold to extract boxes zone. Compared with canny operator, Sobel operator

can detect wispy edge. We select Sobel operator as our model to extract box's edge. We then find contours of boxes edge and use the outline of the moment model to get the contours geometry center. The associated result is depicted in Fig. 2.

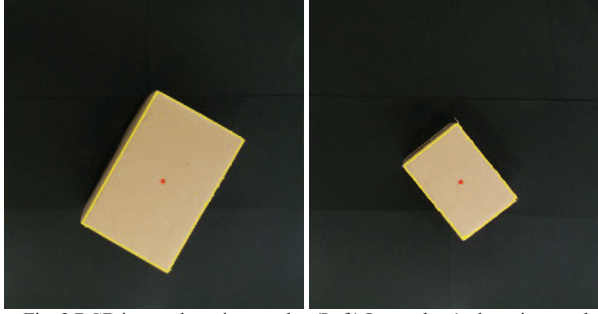


Fig. 2 RGB image detection results. (Left) Larger box's detection result. (Right) Smaller box's detection result. Note that yellow color line is contours of the box's top plane, and red circle point is the corresponding geometry center.

B. Plane segmentation in the Point Cloud

We use Kinect v2 camera as a perception device to get the original data. Origin data are transformed to point cloud by Libfreenect2 and PCL [14].

Point cloud preprocessing, is needed before point cloud plane segmentation. The output point cloud from the Kinect v2 camera contains the X-Y-Z Cartesian coordinates and distance data. Our workplace is constrained to working for a top down view and desktop platform is a fixed zone. What we are concerned about is the object boxes in the desktop. It will make the following algorithm inefficiently by subsequent computational overhead from too many useless data points. Hence, we implement a pass-through filter in the z axis to get rid of points that are too close to the camera. Here the specified range setting is 500–1050 mm. In the same way, we remove other unworkable points in the x and y axis. Although a number of points have removed, there are still some interferential points. We employ a statistical outlier removal method based on Gaussian distribution model to solve this problem. We exclude these undesirable points based on neighborhood statistics by setting the number of neighbors for each point to 30 and related standard deviation of 1.

We want to extract a box grasp plane model from the filtered point cloud data. Random Sample Consensus (RANSAC) [15] algorithm has been widely used in getting 3D model. According to PCL, we can calculate the plane model parameters (1) that best fit into the point cloud.

$$Ax + By + Cz + D = 0 \quad (1)$$

In order to analyze the rotation of box plane, we take approximately segmented plane as cuboid. Here, oriented bounding box in 3D point clouds, which is based on principal component analysis (PCA), is used to estimate main axis of the segmented point cloud plane. We compute the covariance matrix C_i of the 3D point cloud coordinate by (2), where p_i^j is the j^{th} 3D point in point cloud and u^i means the center of point cloud. The N is the number of point cloud points and we calculate the smallest eigenvalue λ and the corresponding eigenvector v of C_i by singular value decomposition (SVD). We consider the corresponding eigenvector of two largest eigenvalues as our box coordinate x and y axis. Meantime, we define z axis through right-hand rule. By this way, we

finally get rotation information of box top plane to point cloud coordinate.

$$C_i = \sum_j^N (p_i^j - u_i)(p_i^j - u_i)^T \quad (2)$$

C. RGB image projection by camera model

In this part, we propose a novel approach to compute box grasp center. Having been validated in [10], method of PCA or MVBB (Minimum Volume Box) cannot get precise pose. Inspired by camera calibration, we first perform RGB camera and IR camera calibration by the approach of [16]. The RGB camera calibration result of intrinsic parameters provides the relationships (3) between image pixel coordinate and camera coordinate. After getting the positions of box's top plane center in the pixel image coordinates in step A, their positions in RGB camera coordinate are easy to be calculated. We suppose that there is one line in RGB camera coordinate's origin to the center point of RGB camera coordinate. Solving the line equation of (4) by using the two points, line model L goes through the two points of O and P can be determined.

$$\begin{bmatrix} su \\ sv \\ 1 \end{bmatrix} = \begin{bmatrix} f_x & 0 & c_x \\ 0 & f_y & c_y \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} X_c \\ Y_c \\ Z_c \end{bmatrix} \quad (3)$$

$$\begin{cases} x = x_0 + at \\ y = y_0 + bt \\ z = z_0 + ct \end{cases} \quad (4)$$

We transform RGB camera line model L into line model L' of IR camera's coordinate by utilizing transformation relationships which are offered by calibration. We calculate the intersecting point of combining plane equation model in step B and L' . The cross point C is the final box's precise center in point cloud coordinate, namely IR camera coordinates. The model philosophy is illustrated in Fig. 3.

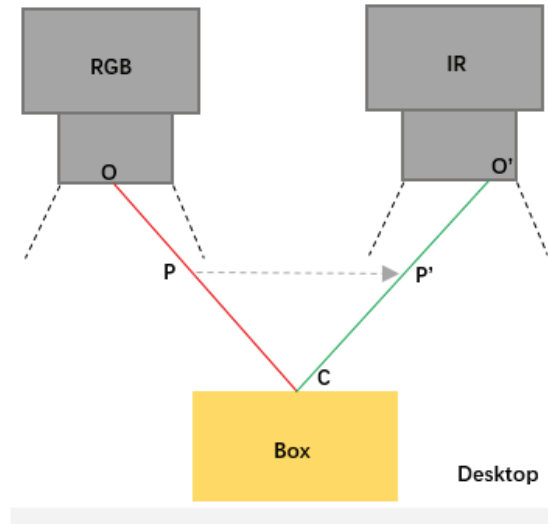


Fig. 3 Projection model in RGB camera and IR camera

IV. EXPERIMENTAL RESULTS

In this section, we present a massive evaluation of the proposed approach on real RGB-D data. And we also weigh the performance of our bin picking setup introduced in below.

Taking cost into account, we set up our bin picking system to verify our algorithm in low cost. Our bin picking system mainly consists of the brain (Industrial Personal Computer), eye (Kinect v2 camera) and hand (ABB IRB1200). Especially, the Kinect v2 depth sensor, is a commercial 3D camera, which is a TOF system with a range of up to 4.5 m and could output 512 x 424 pixel Depth images. Its depth accuracy can up to millimeter according to the measurement of Yang, Lin et al. [17]. We can also capture 1920 x 1080 pixel RGB images while some 3D laser sensors cannot provide. In addition, the robot is equipped with 4 suction cups to grasp boxes. There are also various kinds of size express boxes on the desktop, which is as a workplace. The detailed information can be seen from Fig. 4.

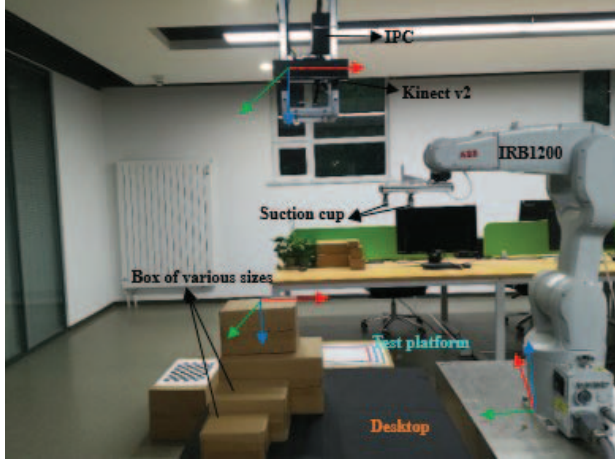


Fig. 4 Design of robotic bin picking system

For improving precision and removing system error, we constructed one template pose before robotic bin picking, which is designed by the routine of feedback open loop control instead of only relying on algorithm direct effect. The final pose measurement and evaluation results are shown in the following steps.

A. Pose Estimation Accuracy

Taking the example of [6], we used different kinds of size express boxes placed on a desktop that was stay horizontal. Owing to its difficulty in estimating quantitatively, we take one chessboard as auxiliary reference and the size of was 9 x 7. First, we calculated the chessboard center pose in the RGB camera coordinate by the method of PnP [18] and regarded chessboard center pose as approximate ground truth. Second, we solved the box's center pose in the point cloud and transformed it into the pose of RGB camera coordinate according to calibration results in section III. Note that the box center and chessboard center were set both same points. Third, we compared the relative pose error between the box center and chessboard center.

We measured 6D pose error by placing boxes in 50 different pose randomly. Fig. 5 depicts the histograms of 6D pose estimation errors of our bin-picking system. Table I shows corresponding average error for each translation and rotation around every axis. The results show the consistent 6D pose estimation results of our system with average translation error is less than 7 mm for x, less than 6 mm for y, and less than 3 mm for z. The rotate angle error is also less than 1°.

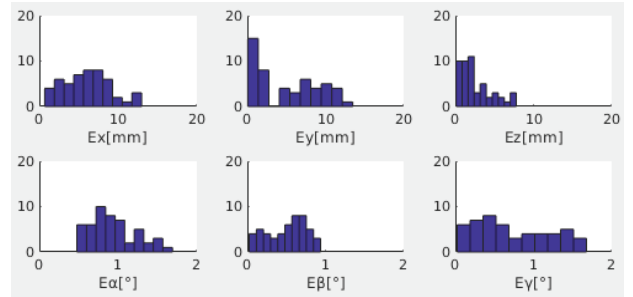


Fig. 5 Histograms of pose estimation errors for each axis of translation (X, Y, Z) and angle (α , β , γ) of rotation around Z, Y, X axis.

TABLE I AVERAGE POSE ESTIMATION ERRORS.

X [mm]	Y [mm]	Z [mm]	α [°]	β [°]	γ [°]
6.18	5.12	2.57	0.94	0.51	0.75

B. Pick and Place Success Rate

We set five different size boxes in desktop by various poses and observed whether the box was placed on the limited area or not. We counted the pick and placement success rate of our bin picking system in this way. The final result was displayed in Table II.

As depicted in Table II, we could find that our system reached a success rate of 98% of 50 trials. The only one test failures were due to strong light illumination and box edge was detected deviating from the center.



Fig. 6 Pick and place experiment. (Top-left) The box's state before picking. (Top-right) The box's state after before picking. (Bottom) Limited error for picking and placing to boxes with different size. If the box is placed inside the corresponding rectangle, we consider the picking up and placing down is successful.

TABLE II PICK AND PLACE SUCCESS RATE.

Total Test	Success	Failure	Success Rate
50	49	1	0.98

C. Processing Time

All tests were implemented on an industrial personal computer with Intel Core i5-5200U CPU 2.20 GHz and 3.8 GB memory. The Linux operating system was used by 64-bit ubuntu16.04 LTS. In addition, the code standard was C++11.

Fig. 7 depicts the comparison of time cost between the RGB to depth image (3D point clouds) region registration method and our camera model projection approach. The processing time of the image based projection approach is 2.97 milliseconds and that of our method is only 0.16 milliseconds. The projection time efficiency has improved with a small quantity. However, as our method has no registration step to construct the relationship in RGB with depth image, we can save 2425.83 milliseconds, which is a remarkable improvement. Especially when the RGB-D camera's image is very high resolution, our method can show the enormous advantage with real-time performance in almost linear time.

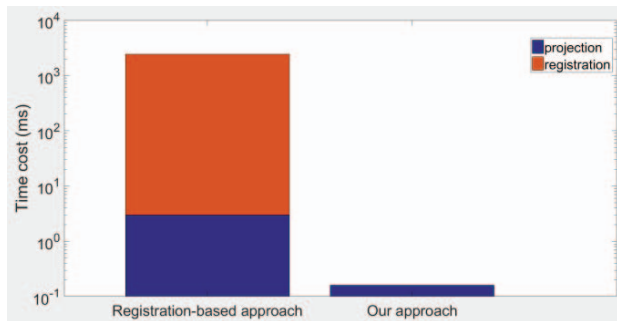


Fig.7 Comparison of time cost of two different methods

Finally, the average time spent on each step of our method is shown in Fig.8. The total time is less than 500 milliseconds, which is real time for bin picking work.

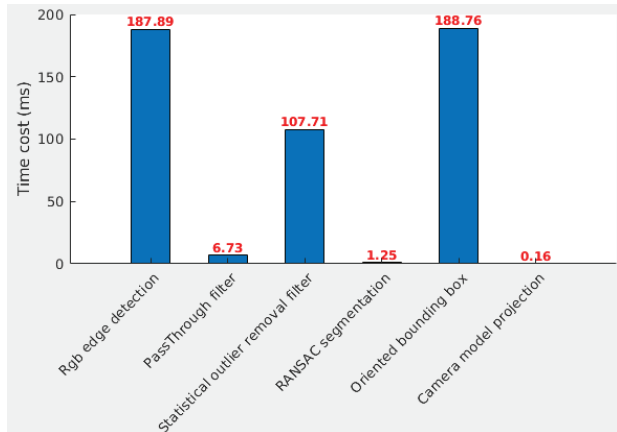


Fig.8 Histograms of time cost evaluation to every step

V. CONCLUSIONS

This paper proposes a novel approach utilizing the principle of the pinhole camera model to obtain object 6D pose for robotic bin picking work from RGB-D data. Instead of only considering 2D images or 3D point cloud, we use pinhole camera model and the geometric relationship to

correlate the point cloud data and RGB pixels for getting highly precise measurement results in position, orientation and picking and placing successfully rate. Firstly, RGB image is converted into HSV channels and the cluttered background is removed by setting proper threshold. After the above process, Sobel operator is used to detect object edge and capture the center of an RGB image. Then, point cloud segmentation is conducted. In order to get foreground objects, point cloud is filtered by pass-through filter and statistical outlier removal filter. Further, the method of the random sample consensus algorithm is employed to extract the object grasp plane. Principal component analysis (PCA) is used to compute object plane rotate direction. Lastly, the homologous binocular calibration principle is adopted to carry on RGB and IR camera calibration of Kinect v2. The RGB image object center is transformed to the IR camera coordinate point by calibration results and pinhole camera model. Simultaneously, compared with the method of single shot of RGB image or point cloud registration, robustness and efficiency of the system is improved a lot. By experiment procedures statistics showing, position's mean error is less than 7 mm for x, less than 6 mm for y, and less than 3 mm for z. And the pose of rotating angle error is also less than 1°. The result validates our approach is excellent compared with other methods.

Our future improvement and refinement will be dedicated to solving the piled and unorderly industry scenarios and realize real time palletizing.

ACKNOWLEDGMENT

This work is supported by the Tianjin Binhai New Area Co-construction Research and Development Platform Science and Technology Project (Grant No. 2017002), and in part by Zhejiang University Students Science and Technology Innovation Activity Plan (New Miao Talent Program) Foundation of 2018.

REFERENCES

- [1] Y. Chen, W. Zhang, G. Zhou, and G. Yan, "A novel building boundary reconstruction method based on lidar data and images," in *International Symposium on Photoelectronic Detection and Imaging 2013: Laser Sensing and Imaging and Applications*, 2013, vol. 8905: International Society for Optics and Photonics, p. 890522.
- [2] Y. Zhang, T. Lu, J. Yang, and H. Kong, "Split and Merge for Accurate Plane Segmentation in RGB-D Images," presented at the 2017 4th IAPR Asian Conference on Pattern Recognition (ACPR), 2017.
- [3] P. J. Besl and N. D. McKay, "Method for registration of 3-D shapes," in *Sensor Fusion IV: Control Paradigms and Data Structures*, 1992, vol. 1611: International Society for Optics and Photonics, pp. 586-607.
- [4] C. Papazov and D. Burschka, "An efficient ransac for 3d object recognition in noisy and occluded scenes," in *Asian Conference on Computer Vision*, 2010: Springer, pp. 135-148.
- [5] B. Drost, M. Ulrich, N. Navab, and S. Ilic, "Model globally, match locally: Efficient and robust 3D object recognition," in *2010 IEEE computer society conference on computer vision and pattern recognition*, 2010: IEEE, pp. 998-1005.
- [6] C. Choi, Y. Taguchi, O. Tuzel, M.-Y. Liu, and S. Ramalingam, "Voting-based pose estimation for robotic assembly using a 3D sensor," in *2012 IEEE International Conference on Robotics and Automation*, 2012: IEEE, pp. 1724-1731.

- [7] R. Vock, A. Dieckmann, S. Ochmann, and R. Klein, "Fast template matching and pose estimation in 3D point clouds," *Computers & Graphics*, vol. 79, pp. 36-45, 2019.
- [8] C. Choi, A. J. Trevor, and H. I. Christensen, "RGB-D edge detection and edge-based registration," in *2013 IEEE/RSJ International Conference on Intelligent Robots and Systems*, 2013: IEEE, pp. 1568-1575.
- [9] P. Chattopadhyay, A. Roy, S. Sural, and J. Mukhopadhyay, "Pose Depth Volume extraction from RGB-D streams for frontal gait recognition," *Journal of Visual Communication and Image Representation*, vol. 25, no. 1, pp. 53-63, 2014.
- [10] S.-H. Lai, V. Lepetit, K. Nishino, and Y. Sato, *Precise Measurement of Cargo Boxes for Gantry Robot Palletization in Large Scale Workspaces Using Low-Cost RGB-D Sensors* (Lecture Notes in Computer Science). 2017.
- [11] W. Kehl, F. Manhardt, F. Tombari, S. Ilic, and N. Navab, "SSD-6D: Making RGB-based 3D detection and 6D pose estimation great again," in *IEEE International Conference on Computer Vision*, 2017.
- [12] C. Wang *et al.*, "DenseFusion: 6D Object Pose Estimation by Iterative Dense Fusion," *arXiv preprint arXiv:1901.04780*, 2019.
- [13] F. Garcia-Lamont, J. Cervantes, A. López, and L. Rodriguez, "Segmentation of images by color features: A survey," *Neurocomputing*, vol. 292, pp. 1-27, 2018.
- [14] R. B. Rusu and S. Cousins, "3D is here: Point Cloud Library (PCL)," in *IEEE International Conference on Robotics & Automation*, 2011.
- [15] R. Schnabel, R. Wahl, and R. Klein, "Efficient RANSAC for point - cloud shape detection," in *Computer graphics forum*, 2007, vol. 26, no. 2: Wiley Online Library, pp. 214-226.
- [16] Z. Zhang, "A flexible new technique for camera calibration," *IEEE Transactions on pattern analysis and machine intelligence*, vol. 22, 2000.
- [17] L. Yang, L. Zhang, H. Dong, A. Alelaiwi, and A. E. Saddik, "Evaluating and Improving the Depth Accuracy of Kinect for Windows v2," *IEEE Sensors Journal*, vol. 15, no. 8, pp. 4275-4285, 2015.
- [18] X.-S. Gao, X.-R. Hou, J. Tang, and H.-F. Cheng, "Complete solution classification for the perspective-three-point problem," *IEEE transactions on pattern analysis and machine intelligence*, vol. 25, no. 8, pp. 930-943, 2003.